

**Universidade Técnica de Lisboa  
Instituto Superior de Economia e Gestão**

Mestrado em Econometria Aplicada e Previsão

2008/2009

# **Modelos com Alteração de Regime: Uma Aplicação Empírica à Taxa de Juro**

Aluno: Patrícia Alexandra Pereira Henriques Varanda

Orientador: Professor Doutor João Carlos Henrique da Costa Nicolau

Júri:

Presidente: Professora Doutora Isabel Maria Dias Proença

Vogal: Professor Doutor João Carlos Henriques da Costa Nicolau

Vogal: Doutor Aníbal Jorge da Costa Cristóvão Caiado

Abril 2011

## *Agradecimentos*

*Quero agradecer em primeiro lugar ao Professor Doutor João Nicolau, orientador desta dissertação, por todos os comentários, sugestões e esclarecimentos dados ao longo da elaboração do presente trabalho, bem como por toda a disponibilidade e dedicação que sempre manifestou ao longo da sua realização.*

*Agradeço ainda ao Professor Doutor Stephen Gray pelo seu contributo para a realização deste trabalho ao disponibilizar o código GAUSS subjacente a algumas das aplicações empíricas apresentadas no seu paper Gray(1996) e que se revelou muito importante como ponto de partida para todo o trabalho que foi desenvolvido nesta dissertação<sup>1</sup>.*

*Todos os erros e omissões devem-se exclusivamente ao autor.*

---

<sup>1</sup> I am also grateful to Professor Stephen Gray for his contribution to this work by providing the GAUSS code used in some of the empirical applications presented in his paper Gray (1996), which has proved to be very important as a starting point for all the work developed in this thesis.

## NOTAÇÃO

Alguma notação e definições *standards* usadas neste trabalho:

- $AR(p)$  significa modelo autoregressivo de ordem  $p$ .
- $y$  é uma série de observações e  $y_t$  constitui uma observação dessa série no momento  $t$ .
- $\Delta y_t = y_t - y_{t-1}$ , ou seja, a variável  $y$  está definida em primeiras diferenças.
- $\Phi_{t-1}$  representa todo o conjunto de informação disponível no momento  $t-1$ , ou seja,  $y_{t-1}, y_{t-2}, y_{t-3}, \dots, y_1$ .
- $\{u_t\}$  e  $\{\varepsilon_t\}$  representam uma sequência de ruído branco, que se caracteriza por possuir média (não condicional) constante e igual a zero; variância (não condicional) constante e covariância (não condicional) nula.  $\{u_t\}$  e  $\{\varepsilon_t\}$  distinguem-se apenas pelo facto de no primeiro caso a variância condicional a  $\Phi_{t-1}$  ser constante, enquanto que, no segundo a variância condicional a  $\Phi_{t-1}$  varia no tempo.
- $T$  representa a última observação da amostra e  $t$  a observação de um determinado momento.
- $h_t$  designa a volatilidade condicional e pode ser entendida como uma medida da magnitude das variações das taxas de juro, não explicadas, no momento  $t$ .
- $Var(\varepsilon_t)$  representa a volatilidade não condicional ou marginal e procura medir a magnitude das variações das taxas de juro, não explicadas, num período de tempo (meses ou anos).
- $k_{z_t}$  é o coeficiente de curtose de  $z_t$  e  $k_{\varepsilon_t}$  é o coeficiente de curtose de  $\varepsilon_t$ .
- $L$  é o operador desfasamento, definido como  $L^k y_t = y_{t-k}$  para  $k = 0, 1, 2, \dots$ .
- $\phi_p(L) = 1 - \phi_1 L - \dots - \phi_p L^p$  é o polinómio autoregressivo em  $L$  de ordem  $p$ .
- i.i.d. significa independente e identicamente distribuído.
- $s_t$  é uma cadeia de *Markov* escondida ou latente porque não é observável.
- $\Phi(\cdot)$  representa a função de distribuição normal cumulativa.

## RESUMO

Este trabalho analisa empiricamente o desempenho de vários modelos lineares na média condicional (ou seja, de um regime) por comparação com modelos que assumem alterações de regime, do tipo *Markov* de ordem um. Em termos de variância condicional acomodam-se várias hipóteses, como a dependência do nível da taxa de juro, de efeitos GARCH ou dos dois factores em simultâneo. A variável em estudo é a taxa de juro nominal a 10 anos da Alemanha. A evidência encontrada foi no sentido de que quer a presença de efeitos GARCH, quer a possibilidade de alteração de regime parecem ser dois factores importantes a ter em conta na modelação da taxa de juro de longo prazo, para efeitos descritivos da série e de previsão *out of sample*.

**Palavras-chave:** modelos não lineares, taxas de juro de longo prazo, previsão, *Markov switching*

**Classificação JEL:** C22, C24, C58, G12

## ABSTRACT

This study empirically examines the performance of conditional mean linear models (single regime) by comparison with first order Markov switching models. Regarding the conditional variance several hypotheses were taken into account, such as the dependence on the level of interest rates, GARCH effects or both. The variable under analysis is the 10 year nominal interest rate in Germany. The evidence supports that either GARCH effects or regime switching seem to be important factors to be taken into account when modelling the long-term interest rate for descriptive and out of sample forecast purposes.

**Key words:** *non linear models; long-term interest rate; forecast; Markov switching*

**JEL classification:** C22, C24, C58, G12

# ÍNDICE

1. Introdução .....	8
2. Modelos de taxas de juro: enquadramento teórico .....	12
2.1. Média condicional .....	14
2.1.1. Média condicional linear .....	14
2.1.2. Média condicional não linear .....	16
2.2. Variância condicional .....	20
2.2.1. Modelos de difusão.....	23
2.2.2. Modelos (G)ARCH lineares .....	24
2.2.3. Modelos GARCH não lineares .....	27
2.3. Modelação da taxa de juro: combinação de modelos .....	29
3. Aplicação empírica .....	33
3.1. Base de dados .....	35
3.2. Estimação, inferência e testes de diagnóstico.....	37
3.2.1. Modelo com o nível da taxa de juro .....	44
3.2.2. Modelo com efeitos GARCH .....	48
3.2.3. Modelo com efeitos GARCH e o nível da taxa de juro .....	53
3.3. Selecção dos modelos e ajustamento <i>in sample</i> .....	58
3.4. Previsão <i>out of sample</i> .....	65
4. Conclusão .....	69
5. Anexo.....	71
Bibliografia.....	76

## LISTA DE GRÁFICOS E DE TABELAS

### Gráficos

Gráfico 1 - <i>Yield</i> do <i>Bund</i> a 10 anos .....	36
Gráfico 2 - Primeiras diferenças ( <i>Bund</i> 10 anos) .....	36
Gráfico 3 - <i>Yield</i> do <i>Bund</i> a 10 anos - valores estimados vs observados .....	62
Gráfico 4 - Média condicional .....	62
Gráfico 5 - Variância condicional .....	62
Gráfico 6 - Variância empírica .....	63
Gráfico 7 - Probabilidade <i>ex-ante</i> .....	64
Gráfico 8 - Probabilidade <i>smoothed</i> .....	64

### Tabelas

Tabela 1 – Estatísticas descritivas de $\Delta r_t$ e $r_t$ .....	37
Tabela 2 – Resultados de estimação do modelo 1.a) e 1.b) .....	46
Tabela 3 – Resultados de estimação do modelo 2.a) e 2.b) .....	49
Tabela 4 – Resultados de estimação do modelo 3.a) e 3.b) .....	54
Tabela 5 – Resultados de estimação do modelo 2.b)* .....	60
Tabela 6 – Previsão <i>out of sample</i> – critérios de avaliação .....	67

# 1. Introdução

A evolução futura das taxas de juro é uma incerteza. Esta questão assume particular pertinência para efeitos de avaliação de instrumentos financeiros e de risco, sendo um assunto com particular interesse para os gestores de *portfolios* com títulos de rendimento fixo, *traders* e actuários. Neste contexto, a capacidade de prever o nível das taxas de juro e a respectiva volatilidade surge como uma questão crítica para os intervenientes dos mercados financeiros.

Existem diferentes taxas de juro associadas a diferentes maturidades e classes de risco. A relação a taxa de juro de activos pertencentes à mesma classe de risco e a respectiva maturidade, num determinado momento de tempo pode ser representada graficamente através da curva de rendimentos. Existem, assim, várias curvas de rendimentos consoante a classe de risco. Esta curva tende assumir formas diversas ao longo do tempo, tais como crescente, decrescente ou *humped*, o que em si mesmo sugere que há um custo associado à detenção de activos com diferentes maturidades, pois caso contrário, a curva assumiria sempre a forma *flat*. O facto de raramente se observar este tipo específico de comportamento indicia que os investidores exigem diferentes taxas de rendibilidade dependendo da maturidade do instrumento que detêm.

Na literatura existe uma grande diversidade de modelos que foram desenvolvidos tendo em vista a modelação da taxa de juro de curto prazo ou, alternativamente, toda a estrutura de prazos. Neste trabalho, a análise centra-se apenas no estudo de um único ponto da curva, mais concretamente, na *yield* na maturidade dos dez anos. Por sua vez, será estudada a classe teoricamente de menor risco, ou seja, a soberana. Pela sua dimensão, liquidez e menor risco de crédito, o mercado de dívida soberana assume uma importância crucial na gestão de investimentos e de risco, de tal forma que nos mercados obrigacionistas é muito comum cotar os *spreads* sobre obrigações soberanas. Por seu turno, na área do euro, as obrigações emitidas pelo Tesouro alemão são, geralmente, consideradas pelos investidores como os activos mais seguros, de tal modo que quando existe um aumento da aversão ao risco estes são o activo de refúgio, beneficiando, frequentemente, do denominado movimento de *flight to quality*. Neste contexto, como a curva soberana alemã é considerada uma referência na Europa para os intervenientes dos mercados financeiros internacionais, sendo frequentemente referida como o *benchmark* de mercado na área do euro, e tendo em



conta o interesse específico em estudar o comportamento da taxa de juro de longo prazo, este trabalho terá como enfoque de análise a *yield* dos títulos de dívida pública *benchmark* alemães a 10 anos. O objectivo é por um lado, explorar alguns modelos no sentido de perceber qual é o que melhor descreve o comportamento histórico da série, captando a sua dinâmica e providenciando o melhor *fit in sample*, e por outro, qual o que mais se adequa para efeitos de previsão *out of sample*. Neste sentido, o estudo apresentado irá focar-se na modelação da média e da variância condicional da taxa de juro a 10 anos, ou seja, da sua rendibilidade e risco, respectivamente.

Tendo por base dados históricos sobre as taxas de juro, existem vários factos estilizados acerca do seu comportamento que são frequentemente referidos na literatura, entre os quais se destacam: 1) a reversão para a média, 2) a leptocurtose da distribuição incondicional da variação da taxa de juro, 3) o fenómeno de *volatility clustering*<sup>2</sup>, 4) a volatilidade mais elevada nas maturidades mais curtas da curva de rendimentos do que nas mais longas, 5) a maior persistência das taxas de longo prazo comparativamente aos prazos mais curtos, e 6) a presença de alguma correlação entre a volatilidade e o nível das taxas de juro. Na literatura também têm surgido referências à existência de diferentes regimes no comportamento das taxas de juro. Por exemplo, Hamilton (1988), Cai (1994), Garcia e Perron (1996), Gray (1996), Bekaert, Hodrick e Marshall (2001) e Ang e Bekaert (2002) estudaram modelos empíricos com alterações de regime nas taxas de juro. Vários autores, como Cecchetti, Lam e Mark (1993) e Garcia (1998) sugerem, ainda, que os modelos de apenas um regime são econometricamente rejeitados a favor dos respectivos modelos com alteração de regime. A característica reversão para a média está relacionada com o facto de se verificar empiricamente que quando as taxas de juro atingem valores muito elevados, subsequentemente tendem a descer em vez de subir e quando alcançam níveis muito baixos acontece o inverso, pelo que as taxas de juro tendem assumir valores dentro de um intervalo delimitado. As taxas de juro não são negativas, logo à partida estão delimitadas inferiormente pelo valor zero e, por outro lado, não é de esperar que tendam para infinito, pelo que de alguma maneira existe também à partida, um limite superior. Em conformidade, seria de esperar, pelo menos teoricamente, que as taxas de juro fossem caracterizadas por um processo de reversão

---

<sup>2</sup> Grandes (reduzidas) variações tendem a ser seguidas por grandes (reduzidas) variações em ambos os sentidos, o que na prática significa que períodos de elevada (baixa) volatilidade tendem a ser seguidos por períodos elevada (baixa) volatilidade. Este fenómeno sugere, assim, que a volatilidade não é constante ao longo do tempo, sendo esta uma característica presente em muitas séries financeiras segundo a evidência empírica.

para a média e, portanto, estacionário. No entanto, na literatura frequentemente quando aplicados os testes de raiz unitária a este tipo de variáveis, os resultados sugerem que a hipótese de raiz unitária não pode ser rejeitada, ou seja, para um determinado nível de significância estatística parece existir evidência de que as taxas de juro não são estacionárias. Esta questão tem sido alvo de alguma controvérsia, já que a evidência empírica de não estacionaridade apontada pelos testes de raiz unitária parece contradizer o que à partida seria de esperar em teoria sobre o comportamento da taxa de juro. Uma das hipóteses que tem sido avançada na literatura para explicar este tipo de resultado, aparentemente controverso, é a possibilidade da taxa de juro ser um processo não linear, e nessas circunstâncias, os testes de raiz unitária têm falta de potência, já que foram desenvolvidos tendo por base a hipótese de linearidade, e daí a sua incapacidade para rejeitar a hipótese nula de raiz unitária. Bekaert *et al.* (2001) e Gray (1996), por exemplo, usando dados para os EUA, sugerem que as taxas de juro tendem a evidenciar um comportamento de passeio aleatório apenas quando as taxas de juro se encontram em níveis reduzidos, enquanto que quando as taxas de juro assumem valores mais elevados se observa um comportamento de reversão para a média. Tendo em conta estes aspectos e a variável em estudo, os modelos não lineares surgiram naturalmente como uma hipótese de investigação a considerar. Assim, neste trabalho irão ser abordados os modelos não lineares, a par dos lineares, com o intuito de avaliar e comparar as suas capacidades descritivas e de previsão a curto prazo, procurando tirar ilações sobre a sua potencial relevância.

No que se refere aos outros factos estilizados das taxas de juro acima referidos e que dizem respeito ao comportamento da volatilidade é de notar que surgiram vários modelos na literatura com o intuito de captar o melhor possível a sua dinâmica. Entre os mais populares destacam-se os pertencentes à classe **(G)ARCH** (*Generalized AutoRegressive Conditionally Heteroscedastic*) e os modelos de **difusão**. Estes dois tipos modelos têm a capacidade de captar características como a *volatility clustering* e a persistência, contudo, distinguem-se entre si pelo facto de no primeiro caso se assumir que as trajectórias decorrem de processos estocásticos definidos em tempo discreto, enquanto que no segundo caso se admite que decorrem de processos estocásticos definidos em tempo contínuo.

No contexto da modelação das taxas de juro existe uma grande diversidade de modelos possíveis, em particular na classe dos modelos não lineares. Neste trabalho optou-se por seguir a metodologia desenvolvida por Gray (1996), por ser uma

abordagem genérica, que engloba vários modelos de taxas de juro, e também por este ter sido pioneiro na introdução dos modelos GARCH no contexto não linear. Tendo em conta que este tipo de modelos se revelam particularmente úteis para captar características como a persistência e dado que a evidência empírica sugere que os prazos mais longos da curva de rendimentos tipicamente revelam uma elevada persistência, considerou-se oportuno estudar este tipo de abordagem. Os modelos não lineares explorados por este autor inserem-se na classe de *Markov switching* (MSW), que se caracterizam pelo facto de a alteração de regime ser determinada por um processo não observável.

Os resultados obtidos neste trabalho parecem indicar que modelos analisados que incorporam a possibilidade de alteração de regime captam melhor o comportamento da taxa de juro de longo prazo comparativamente aos modelos com apenas um regime. Adicionalmente, existem indícios de elevada persistência da taxa de juro de longo prazo, em particular da sua variância, pelo que é necessário incorporar efeitos GARCH para capturar adequadamente a sua dinâmica. A evidência empírica sugere que o modelo com alteração de regime e com efeitos GARCH face ao modelo de um regime produz, por um lado, um melhor ajustamento aos dados históricos, e por outro, uma melhor capacidade de previsão *out of sample* para o horizonte de um a três meses.

A forma como o presente trabalho se encontra organizado é a seguinte. No **capítulo 2** apresenta-se um enquadramento teórico sobre algumas formulações lineares e não lineares usadas na literatura na modelação das taxas de juro. Assim, em primeiro lugar, será efectuada uma breve introdução sobre a modelação da média condicional num contexto linear a que se seguirá uma pequena exposição dos principais aspectos metodológicos a ter em conta no quadro de não linearidade. Posteriormente, apresenta-se uma referência genérica a alguns dos modelos mais relevantes na modelação da variância condicional, como é o caso dos modelos da classe GARCH e de difusão. O capítulo termina com uma referência a alguns modelos, que combinam as formulações de média e da variância condicional abordados anteriormente, com vista a captar adequadamente alguns factos estilizados sobre as taxas de juro. No **capítulo 3** discutem-se questões relacionadas com a estimação e são apresentados os resultados empíricos de alguns modelos lineares e não lineares, seguindo de perto a abordagem de Gray (1996). Por fim, o **capítulo 4** sintetiza as principais conclusões deste trabalho.

## 2. Modelos de taxas de juro: enquadramento teórico

Neste trabalho não existe a pretensão de modelar toda a curva de rendimentos, mas somente a dinâmica de um ponto específico da curva nas maturidades mais longas, mais concretamente, a *yield* de obrigações do Tesouro a 10 anos, que é considerada como uma maturidade de referência nos mercados obrigacionistas. Ao colocar o enfoque num único ponto pretende-se, por um lado, compreender como se tem caracterizado o comportamento das taxas de juro de longo prazo ao longo do tempo e se os resultados encontrados são análogos aos apresentados na literatura para os prazos mais curtos da curva de rendimentos. Por outro lado, tendo em conta que um dos objectivos principais é avaliar a capacidade de previsão da taxa de juro a 10 anos, comparando a *performance* de modelos lineares e não lineares, foi dada preferência ao uso de modelos univariados em detrimento dos multivariados. Esta opção é menos exigente do ponto de vista de complexidade computacional, permitindo manter os modelos simples e flexíveis o suficiente de modo a capturar diversas situações, em particular, ao nível da dinâmica da volatilidade. De notar ainda, que a teoria de modelos não lineares multivariados em séries temporais ainda é relativamente recente, existindo mesmo nalguns casos resultados conhecidos apenas para alguns modelos específicos, e noutros casos a teoria ainda se encontra em fase de desenvolvimento e, nalguns aspectos, mesmo por desenvolver.

Os modelos de taxas de juro devem procurar captar alguns factos estilizados sobre o comportamento empírico das taxas de juro já mencionados anteriormente, entre os quais se destaca a reversão para a média e o facto da distribuição não condicional da variação das taxas de juro ser leptocúrtica. De entre os modelos existentes que visam modelar a taxa de juro de curto prazo, uma forma habitual de captar a característica de reversão para a média consiste em definir a variação da taxa de juro num determinado momento, como uma função linear do nível da taxa de juro do período anterior. Por sua vez, Engle (1982) mostrou que uma possível causa para a distribuição não condicional ser leptocúrtica é a presença de heterocedasticidade condicional. Na literatura há duas formas comuns para modelar a heterocedasticidade condicional. Uma forma possível, e que é frequentemente utilizada nos denominados modelos de difusão, consiste em modelar a variância condicional da variação da taxa de juro como uma função do nível

da taxa de juro<sup>3</sup>. A outra alternativa insere-se no contexto dos modelos da classe (G)ARCH, que foram introduzidos por Engle (1982) e Bollerslev (1986). Gray (1996) foi mais longe, adoptando uma abordagem mais genérica, com o argumento de que em determinadas aplicações empíricas estas duas formas de modelar a heterocedasticidade condicional poderiam não ser suficientes para captar o grau de leptocurtose presente nos dados. Neste sentido, Gray considerou simultaneamente o efeito dos choques desfasados ao quadrado e a variância no período anterior (modelos GARCH), bem como a hipótese adicional da variância depender de efeitos GARCH e do nível da taxa de juro no período anterior, em contexto linear e não linear.

A análise empírica abordada no **capítulo 3** irá ser efectuada tendo por base modelos lineares e não lineares, seguindo de perto abordagem de Gray(1996). Os modelos não lineares ainda são relativamente recentes na literatura, contudo, existe já uma grande diversidade de formulações propostas, sendo que teoricamente há uma infinidade de possibilidades em termos de especificações não lineares (ver Tong (1990) e Granger e Teräsvirta (1993)). Este tipo de modelos permite a modelação de variáveis admitindo diferentes regimes, de tal forma, que acomodam a possibilidade de a variável evidenciar diferentes dinâmicas consoante o regime em que se encontra num determinado momento do tempo. Os modelos do tipo linear assumem uma estrutura fixa e única para a média e para a variância condicional em todo o período amostral. Por seu turno, nos modelos do tipo não linear esta hipótese é relaxada, e admite-se a possibilidade de os parâmetros da média e da variância variarem no tempo em função da alteração de regime. Assim, certas propriedades, como a média, a variância ou mesmo a autocorrelação podem ser diferentes em função do regime. Os modelos com alteração de regime são mais flexíveis, com capacidade de incorporar não linearidades significativas, mas simultaneamente mais complexos e sem a teoria completamente desenvolvida. A pertinência deste tipo de modelos para o caso das taxas de juro advém do facto de historicamente existirem períodos em que se observaram taxas de juro elevadas e outros intervalos de tempo em que as taxas de juro tenderam assumir valores mais reduzidos, o que aparentemente poderá sugerir a existência de diferentes regimes ao longo do tempo. Este tipo de considerações também parece extensível à variância exibida por este tipo de variáveis. Em particular, entre 1979 e 1982, alterações

---

<sup>3</sup> O pressuposto de a volatilidade ser representada como uma função do nível da taxa de juro é a abordagem seguida pelos modelos de difusão unifactoriais que procuram modelizar a dinâmica da taxa de juro de curto prazo a partir de uma equação diferencial estocástica. O modelo desenvolvido por Cox, Ingersoll e Ross (1985) insere-se neste tipo de abordagem.

substâncias em termos de condução da política monetária, levadas a efeito pela Reserva Federal, actuação que ficou conhecida como *Federal Reserve Experiment*, conduziram um acréscimo substancial de volatilidade das taxas de juro nos EUA. Este tipo de comportamento motivou aplicação de modelos com alteração de regime no contexto da modelação das taxas de juro, destacando-se a este respeito o trabalho desenvolvido por Hamilton (1988) e Cai (1994). O aumento substancial da volatilidade verificou-se ainda em outros períodos marcantes da história do EUA, como as crises petrolíferas, o *crash* do mercado accionista em Outubro 1987 e períodos de guerra.

De seguida efectua-se uma caracterização genérica de algumas formulações usadas na modelação da **média condicional**, introduzindo o tema com os modelos **lineares** e extendendo, posteriormente, a análise para o contexto **não linear**. Posteriormente resume-se genericamente alguns modelos lineares e não lineares usados frequentemente na literatura na modelação da **variância**. Por fim, são apresentados alguns modelos de taxa de juro que misturam as formulações descritas anteriormente para a média e para a volatilidade, num contexto de **MSW**, que constitui a categoria de modelos não lineares estudados em detalhe na parte empírica deste trabalho.

## 2.1. Média condicional

### 2.1.1. Média condicional linear

A classe mais genérica de modelos lineares univariados introduzidos por Box e Jenkins (1970) é conhecida na literatura por **ARMA( $p, q$ )** (*Autoregressive Moving Average*) e caracteriza-se por combinar processos **AR( $p$ )** (*Autoregressive*) e **MA( $q$ )** (*Moving Average*) que podem ser entendidos como casos particulares, dessa formulação mais geral. Este tipo de modelos são muito populares na literatura, já que são modelos lineares simples e, por vezes, revelam-se úteis para efeitos de descrição e previsão de determinadas séries. O objectivo principal na modelação da média condicional com recurso a modelos pertencentes à classe ARMA é o de definir um modelo parcimonioso, em termos de parâmetros, que apresente boas propriedades estatísticas e descreva bem a série em estudo. A escolha por uma determinada especificação em detrimento de outra prende-se com as características da série em estudo, em particular, com o nível da autocorrelação entre as observações. Nos processos do tipo AR a questão prende-se com

a evidência de autocorrelação, ou seja, este tipo de modelo pode ser apropriado se os valores passados da série  $y$  podem explicar parcialmente o movimento presente de  $y$ . A formulação AR tem-se revelado particularmente popular dada a sua capacidade de reproduzir razoavelmente a dinâmica de muitas séries financeiras e económicas. Por sua vez, a especificação MA é mais adequado para modelar fenómenos de memória muito curta, pois prova-se que neste tipo de modelos a autocorrelação de  $y$  desaparece muito rapidamente. Tendo em consideração que a evidência empírica sugere que normalmente nas séries financeiras a média condicional quando modelada através de um  $AR(p)$ ,  $p$  não é muito elevado e que nos modelos de taxas de juro este tipo de formulação revela-se geralmente mais adequada do que a MA, sendo, por isso, a mais frequentemente usada<sup>4</sup>, neste tópico será apenas dado enfoque aos modelos AR, cuja especificação será tida em conta na aplicação empírica, no **capítulo 3**.

Com efeito, num enquadramento  $AR(p)$ , seja  $y_t$  a variável em estudo e observada para  $t=1,2,\dots,T$  e admitindo que os valores pré-amostrais  $y_0, y_{-1}, \dots, y_{1-p}$  estão disponíveis sempre que necessário, uma forma comum de representar a componente explicativa é assumir que  $y_t$  pode ser expresso como uma combinação linear dos seus  $p$  valores desfasados e, eventualmente, de uma constante. Ou seja,

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + u_t, \quad t = 1, 2, \dots, T \quad (1)$$

onde  $\phi_0, \phi_1, \dots, \phi_p$  são parâmetros desconhecidos a ser estimados e  $\{u_t\}$  é um ruído branco que verifica a condição  $E(u_t | \Phi_{t-1}) = 0$  (sendo que  $\Phi_{t-1}$  representa todo o conjunto de informação disponível no momento  $t-1$ ). Tal como já mencionado, para a modelação da média condicional de muitas séries financeiras é provável que a informação recente seja suficiente e, portanto,  $p$  seja relativamente pequeno. Usando o operador desfasamento,  $L$ , pode-se reescrever a equação (1) acima apresentada como:

$$\phi_p(L) y_t = \phi_0 + u_t, \quad t = 1, 2, \dots, T \quad (2)$$

onde  $\phi_p(L) = 1 - \phi_1 L - \dots - \phi_p L^p$  representa o polinómio autoregressivo em  $L$  de ordem  $p$ .

---

<sup>4</sup> De notar, ainda, que na modelação com processos ARMA há o risco de parametrização redundante (ver exemplos apresentados em Hamilton (1994)).

A estacionaridade em covariância de uma série,  $y_t$ , gerada por um modelo  $AR(p)$  é determinada pelos parâmetros  $\phi_1, \dots, \phi_p$ . A equação característica neste caso é dada por:

$$1 - \phi_1 z - \dots - \phi_p z^p = 0. \quad (3)$$

A condição de estacionaridade em covariância exige simplesmente que todas as  $p$  soluções, ou raízes, desta equação característica sejam superiores a um. Como as soluções podem ser números complexos, na prática esta condição implica que todas as raízes estão fora do círculo unitário.

Assumindo que  $y_t$  é estacionário em covariância e que  $E[y_t] = \mu$  e  $E[u_t] = 0 \quad \forall t = 1, \dots, T$ , obtém-se facilmente que a média não condicional é dada por:

$$\mu = \frac{\phi_0}{1 - (\phi_1 + \dots + \phi_p)}. \quad (4)$$

De notar que apesar de num processo estacionário a média não condicional ou marginal ser constante, o mesmo não acontece no caso da média condicional, que é variável ao longo do tempo, e neste caso dada por:

$$E[y_t | \Phi_{t-1}] = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p}, \quad t = 1, 2, \dots, T. \quad (5)$$

### 2.1.2. Média condicional não linear

Os modelos com alteração de regime podem ser subdivididos em duas categorias, que diferem entre si na forma como o regime evolui ao longo do tempo.

A **primeira categoria** de modelos caracteriza-se pelo facto de assumir que o regime que ocorre no momento  $t$  ser determinado por uma variável observável ( $q_t$ ). Em conformidade, em cada momento de tempo conhece-se com certeza o regime em que o processo se encontra porque a mudança de regime depende de uma variável observável. Nesta categoria de modelos insere-se o modelo limiar autoregressivo (*Threshold Autoregressive* – **TAR**) proposto inicialmente por Tong (1978) e Tong e Lim (1980), e



discutido em profundidade em Tong (1990). O modelo TAR assume que o regime é determinado pelo valor da variável  $q_t$  relativamente a um certo valor limiar (*threshold value*),  $c$ . Com efeito, o modelo TAR geral (considerando por simplificação de notação, apenas variáveis endógenas) que consubstancia a existência de  $k$  regimes e  $p$  defasamentos, para o processo  $y$ , pode ser representado da seguinte forma<sup>5</sup>:

$$y_t = \begin{cases} \phi_{01} + \phi_{11}y_{t-1} + \dots + \phi_{p1}y_{t-p} + u_t & \text{se } q_t < c_1 \\ \dots & \dots \\ \phi_{0k} + \phi_{1k}y_{t-1} + \dots + \phi_{pk}y_{t-p} + u_t & \text{se } q_t > c_k \end{cases} \quad (6)$$

onde  $\phi_{ji}$ , com  $j = 0, \dots, p$  e  $i = 1, \dots, k$  são parâmetros a estimar e  $\{u_t\}$  é um ruído branco.

A ideia base deste tipo de modelos consiste em assumir processos lineares específicos a cada regime assumindo-se, assim, que  $y$  apenas se comporta de forma linear em subintervalos do seu domínio. Assim, na representação do processo  $y$  acima apresentada pode-se dizer que a especificação linear  $\phi_{01} + \phi_{11}y_{t-1} + \dots + \phi_{p1}y_{t-p} + u_t$  deverá caracterizar adequadamente o processo  $y$  quando  $q_t$  assume um valor inferior ao limiar  $c_1$ , ou seja, no regime 1. Por seu turno, no regime  $k$ , quando a variável limiar assume um valor superior a  $c_k$ , o processo será melhor representado por uma especificação linear alternativa, que neste caso seria  $\phi_{0k} + \phi_{1k}y_{t-1} + \dots + \phi_{pk}y_{t-p} + u_t$ .

No caso em que o regime é determinado pela série em estudo ( $y$ ), o modelo designa-se por *Self-Exciting TAR (SETAR)*. Esta designação advém do facto de neste caso a variável que determina a mudança de regime,  $q_t$ , corresponder à própria variável em estudo desfasada, ou seja,  $q_t = y_{t-d}$  (sendo  $d$  um valor inteiro positivo). No modelo SETAR assume-se que a fronteira entre os regimes é dada por um valor específico,  $c_i$  com  $i = 1, \dots, k$ , para a variável *threshold*,  $y_{t-d}$ , existindo uma mudança imediata de regime quando o valor *threshold* é atingido. Existe, contudo, a possibilidade de a transição entre regimes ocorrer de forma suave. Os modelos que acomodam esta hipótese designam-se de *Smooth Transition Autoregressive (STAR)* e caracterizam-se por incorporar uma função de transição entre regimes, que permite suavizar a transição

---

<sup>5</sup> Quer nos modelos da primeira categoria quer da segunda, a exposição será efectuada assumindo que existe apenas uma variável exógena, que é o termo constante, tendo em consideração que é nesse enquadramento que será desenvolvida a parte empírica apresentada no **capítulo 3**. Contudo, sob determinados pressupostos, os modelos apresentados podem ser generalizados para o caso em que existem outras variáveis exógenas.

entre regimes. Estes modelos foram introduzidos na literatura por Chan e Tong (1986) e, posteriormente, popularizados por Granger e Teräsvirta (1993) e Teräsvirta (1994)<sup>6</sup>.

Já os modelos pertencentes à **segunda categoria**, usualmente são conhecidos por **MSW** e caracterizam-se por assumirem que o regime num certo momento  $t$  não pode ser observado porque é determinado por um processo não observável, que será designado doravante por  $s_t$ .  $s_t$  é o indicador de regime latente e segue uma cadeia de *Markov* com um espaço de estados finito discreto, ou seja,  $s_t = 1, 2, \dots, k$ . Em conformidade, neste tipo de modelos não se sabe com certeza qual o regime que ocorreu num certo momento, existindo apenas uma probabilidade associada a tal acontecimento.

Os modelos mais populares desta categoria são os modelos de **MSW**, introduzidos por Hamilton (1988, 1989, 1990), e segundo os quais  $s_t$  segue uma **cadeia de Markov homogénea<sup>7</sup> de ordem um**, ou seja, o regime actual  $s_t$  depende apenas do regime do período anterior  $s_{t-1}$ , sendo independente de  $s_{t-2}, s_{t-3}, \dots, s_0$  e de  $\Phi_{t-1}$  dado  $s_{t-1}$ . Mais formalmente  $\Pr(s_t = i | s_{t-1} = j, \Phi_{t-1}) = \Pr(s_t = i | s_{t-1} = j) = p_{ji}$ . Em termos gerais, o processo  $y$  pode ser representado da seguinte forma:

$$y_t = \begin{cases} \phi_{01} + \phi_{11}y_{t-1} + \dots + \phi_{p1}y_{t-p} + u_t & \text{se } s_t = 1 \\ \dots & \dots \\ \phi_{0k} + \phi_{1k}y_{t-1} + \dots + \phi_{pk}y_{t-p} + u_t & \text{se } s_t = k \end{cases} \quad (7)$$

onde  $k$  representa o número de regimes,  $p$  o número de defasamentos das variáveis,  $\phi_{ji}$ , com  $j = 0, \dots, p$  e  $i = 1, \dots, k$  são parâmetros a estimar e  $\{u_t\}$  é um ruído branco. Para o modelo ficar completo é ainda necessário definir a matriz de probabilidades de transição entre regimes,  $M$ , ou seja:

$$M = \begin{pmatrix} p_{11} & \dots & p_{1k} \\ \vdots & \ddots & \vdots \\ p_{k1} & \dots & p_{kk} \end{pmatrix} \quad (8)$$

<sup>6</sup> Para uma revisão mais profunda dos modelos STAR ver Teräsvirta (1998).

<sup>7</sup> Significa que as probabilidades de transição de *Markov* são constantes e, portanto, a matriz  $M$ , não depende de  $t$ .

sendo que  $p_{ji}$  representa a probabilidade da cadeia de *Markov* passar do regime  $j$  em  $t-1$  para o regime  $i$  no momento  $t$ . Neste contexto, uma cadeia de *Markov* é completamente caracterizada pela matriz de probabilidades de transição,  $M$ , que tem de satisfazer duas condições:

1)  $p_{ji} \geq 0$  para  $j, i = 1, \dots, k$ , que traduz o facto de como  $p_{ji}$  representa uma probabilidade, para estar correctamente definida, não pode assumir valores negativos;

2)  $\sum_{i=1}^k p_{ji} = 1$  para  $j = 1, \dots, k$ , ou seja, a probabilidade de estar no regime  $j$  em

$t-1$  e manter-se nesse regime em  $t$ , acrescida do somatório das probabilidades de estar no regime  $j$  em  $t-1$  e mudar para qualquer um dos outros  $k-1$  regimes em  $t$ , naturalmente tem de somar um.

Ainda dentro desta abordagem e sob certas condições, é possível determinar a probabilidade de estar num determinado regime no longo prazo independentemente do estado inicial, ou seja,  $P(s_t = j)$  para  $j = 1, 2, \dots, k$ , também designada por probabilidade não condicional ou estacionária. Formalmente, a questão resume-se a determinar  $\lim_{n \rightarrow \infty} P(s_{t+n} = j | s_t = i) = P(s_t = j) = \pi_j \geq 0$ , para qualquer  $i$ . Nem todas as cadeias de *Markov* admitem estas probabilidades limite, mas sob certas condições, pode-se provar (veja-se Karlin *et al.* (1998)) que o vector (linha) das **probabilidades estacionárias** é dado por  $\pi = (\pi_1 \ \pi_2 \ \dots \ \pi_k)$ , onde  $\pi_j$  satisfaz duas equações: 1)

$\pi = \pi M$  e 2)  $\sum_{j=1}^k \pi_j = 1$ . Usando a teoria ergódica das cadeias de *Markov* é possível

mostrar (para uma derivação explícita deste resultado ver Hamilton (1994)) que as probabilidades incondicionais no modelo de **MSW** com dois regimes são dadas por:

$$P(s_t = 1) = \frac{1 - p_{22}}{1 - p_{11} - p_{22}} \quad , \quad P(s_t = 2) = \frac{1 - p_{11}}{1 - p_{11} - p_{22}} . \quad (9)$$

As condições de estacionaridade para modelos do tipo **MSW** com dois regimes foram discutidas por Holst *et al.* (1994). Stelzer (2009) também estabeleceu algumas condições de estacionaridade para alguns modelos mais gerais, com  $k$  regimes, em contexto de modelos com homocedasticidade condicional e Bauwens *et al.* (2006) no caso de heterocedasticidade condicional.

No contexto não linear, e mantendo a hipótese de  $s_t$  seguir uma cadeia de *Markov* de ordem um, é possível ainda assumir uma abordagem alternativa relativamente à matriz de probabilidades de transição. Com efeito, é possível assumir que  $s_t$  depende de  $\Phi_{t-1}$  e, portanto, neste contexto  $p_{ji}$  deve ser substituída por  $p_{ji_t}$  onde  $p_{ji_t} = \Pr(s_t = i | s_{t-1} = j, \Phi_{t-1})$ . Este enquadramento é usualmente conhecido por **MSW com probabilidades de transição que variam no tempo**. Dentro desta abordagem uma possibilidade consiste em definir aquela probabilidade,  $p_{ji_t}$ , como uma função da variável  $r_{t-1}$ , que, seguindo Gray (1996), será o caso em apreço na parte empírica<sup>8</sup>. Esta hipótese permite estabelecer uma relação entre o nível de taxas de juro e a probabilidade de transitar de um determinado regime para outro. Existem várias alternativas para modelar  $p_{ji_t}$ , sendo que os modelos de variável dependente binária *logit* ou *probit* surgem na literatura como os mais comuns para o caso de modelos com dois regimes.

## 2.2. Variância condicional

Nos **modelos lineares**, e de forma genérica nos **não lineares**<sup>9</sup>, qualquer série  $y_t$  pode ser interpretada como sendo a soma de duas componentes: uma previsível, na medida em que pode ser prevista usando o conhecimento que temos do passado,  $\Phi_{t-1}$ , e outra imprevisível e denotada por  $u_t$ . Ou seja, a série  $y_t$  pode ser decomposta como:

$$y_t = E[y_t | \Phi_{t-1}] + u_t \quad (10)$$

onde  $E[.]$  representa o operador de expectativa condicional. De notar que  $\{u_t\}$  é ruído branco e que verifica a propriedade  $E[u_t | \Phi_{t-1}] = 0$ . No subcapítulo anterior, **2.1.**, foi efectuado um enquadramento relativamente à modelação da média condicional,  $E[y_t | \Phi_{t-1}]$ , tendo já sido referido nesse contexto, que o processo  $y_t$  poderia ser adequadamente caracterizado por uma combinação linear de  $p$  valores desfasados (modelo autoregressivo de ordem  $p$ ,  $AR(p)$ ), ou seja,

<sup>8</sup> Para exemplos adicionais da abordagem MSW com probabilidades de transição que variam no tempo ver Diebold *et al.* (1994) e Filardo (1994).

<sup>9</sup> No sentido em que, em cada regime se assume um comportamento linear em termos de média condicional.

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + u_t \quad t = 1, \dots, T \quad (11)$$

onde  $\phi_1, \dots, \phi_p$  são parâmetros desconhecidos a ser estimar. Contudo, nesse enquadramento assumia-se, simplesmente, que a parte imprevisível,  $u_t$ , satisfazia as propriedades de ruído branco. Nomeadamente, assumia-se homocedasticidade não condicional e condicional de  $u_t$ , ou seja,  $E[u_t^2] = E[u_t^2 | \Phi_{t-1}] = \sigma^2$  para qualquer  $t$ . Neste subcapítulo, esta hipótese vai ser colocada de parte e vai ser considerada a possibilidade de a variância condicional variar no tempo<sup>10</sup>. Com efeito, serão apresentados vários modelos que admitem esta possibilidade, permitindo, assim, modelar a volatilidade<sup>11</sup>, que é uma variável fundamental na gestão do risco nos mercados. Adicionalmente, se um processo,  $y$ , exibe heterocedasticidade condicional, a estimação conjunta dos parâmetros da média e da variância, permite estimar de forma mais eficiente os parâmetros da média e, simultaneamente, obter intervalos de confiança mais correctos para  $y$ .

De forma genérica, sendo  $\varepsilon_t$  o choque da equação, que se caracteriza por ser condicionalmente heterocedástico, pode-se considerar a seguinte representação:

$$\varepsilon_t = z_t \sqrt{h_t} \quad (12)$$

onde  $z_t$  é independente e identicamente distribuído (i.i.d.), com média zero e variância unitária;  $h_t$  é uma função não negativa e representa a variância condicional de  $\varepsilon_t$ , ou seja,  $E(\varepsilon_t^2 | \Phi_{t-1}) = h_t$ . Se se assumir que  $z_t$  tem distribuição normal,  $\varepsilon_t$  terá uma distribuição condicional em  $\Phi_{t-1}$  normal com média zero e variância  $h_t$ , ou seja,  $\varepsilon_t | \Phi_{t-1} \sim N(0, h_t)$ . Pode-se demonstrar<sup>12</sup> que  $k_{\varepsilon_t} \geq k_{z_t}$ , logo como se assume que  $z_t \sim N(0,1)$ , a distribuição de  $\varepsilon_t$  tem caudas mais pesadas do que a distribuição de  $z_t$ , apresentando, por isso, uma distribuição marginal leptocúrtica.

---

<sup>10</sup> Em termos de notação e por uma questão de clareza, optou-se por designar o choque ou erro de cada equação, por  $u_t$  quando se assume homocedasticidade condicional e não condicional, e alternativamente por  $\varepsilon_t$  quando se admite a possibilidade de heterocedasticidade condicional, continuando-se, no entanto, assumir homocedasticidade não condicional.

<sup>11</sup> Neste trabalho, a volatilidade condicional será identificada como a variância condicional, ou seja,  $h_t$ .

<sup>12</sup> Ver Engle (1982) a demonstração desta propriedade no contexto de modelos ARCH(1).

De notar que quando a variável residual do modelo é heterocedástica, essa característica é transmitida à série em estudo, neste caso,  $y_t$ . Sendo  $y_t$  definido tal como na formulação (11), a variância condicional de  $y_t$ , é dada por:

$$Var(y_t | \Phi_{t-1}) = E((y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \dots - \phi_p y_{t-p})^2 | \Phi_{t-1}) = Var(\varepsilon_t | \Phi_{t-1}) = h_t. \quad (13)$$

Na literatura existem vários modelos desenvolvidos com o objectivo de modelar a volatilidade, sendo os do tipo **(G)ARCH**, introduzidos por Engle (1982) e Bollerslev (1986), os mais populares. Em particular, este tipo de modelos assumiu uma relevância significativa na modelação de séries financeiras dada a sua capacidade para descrever alguns factos estilizados que estão geralmente, associados a este tipo de séries, como a característica de *volatility clustering* e de caudas pesadas<sup>13</sup>. Por outro lado, a evidência empírica sugere também que a volatilidade tende a ser maior quando a taxa de juro de curto prazo é mais elevada. Este efeito do nível é normalmente captado pelos modelos de **difusão**. Assim, de seguida irá procurar-se resumir genericamente estas duas formas alternativas, comuns na literatura, de modelar a heterocedasticidade. O objectivo deste subcapítulo não é o de abordar de forma ampla toda a literatura existente sobre vários modelos de difusão e do tipo (G)ARCH, ou mesmo outras categorias de modelos que também procuram modelar a volatilidade, como por exemplo os que pertencem à classe da volatilidade estocástica (*Stochastic Volatility* - SV), cujo modelo básico foi introduzido por Taylor (1986). Ao invés será efectuada apenas uma breve referência aos modelos relevantes na aplicação empírica apresentada no **capítulo 3**. Com efeito, inicialmente irá apresentar-se uma síntese da formulação subjacente aos modelos de difusão, a que se seguirá uma introdução teórica do modelos mais simples do tipo GARCH (GARCH(1,1)), em contexto linear. Por fim, será efectuada um enquadramento desta última categoria de modelos assumindo um cenário de não linearidade em contexto de **MSW**, seguindo o procedimento desenvolvido por Gray (1996).

---

<sup>13</sup> Os modelos de heterocedasticidade condicional implicam, sob certas condições, distribuições marginais leptocúrticas, revelando-se, por isso, adequados para modelar séries com este tipo de distribuição.

### 2.2.1. Modelos de difusão

Neste enquadramento, assume-se que a taxa de juro é um processo estocástico em tempo contínuo que pode ser bem representado por um processo de difusão do tipo:

$$dr_t = (a + br_t)dt + \sigma r_t^\gamma dW_t \quad (14)$$

onde  $r$  é a taxa de juro *spot* instantânea e  $W$  é o chamado processo de Wiener. Esta equação diferencial estocástica, da forma apresentada, engloba vários modelos usados na literatura, tal como referem Chan, Karolyi, Longstaff e Sanders (1992). Em particular, quando  $\gamma = 0.5$ , a equação descreve o processo CIR<sup>14</sup> que é um modelo muito popular em matemática financeira no âmbito da modelação das taxas de juro.

Em termos genéricos, a equação diferencial estocástica discretizada<sup>15</sup> de acordo com o esquema de Euler (ver Chan *et al.*) conduz à seguinte representação:

$$\begin{aligned} \Delta r_t &= r_t - r_{t-1} = a + br_{t-1} + \varepsilon_t \\ \varepsilon_t &= z_t \sqrt{h_t} \\ h_t &= \sigma^2 r_{t-1}^{2\gamma} \end{aligned} \quad (15)$$

onde  $E[\varepsilon_t | \Phi_{t-1}] = 0$  e  $E[\varepsilon_t^2 | \Phi_{t-1}] = \sigma^2 r_{t-1}^{2\gamma}$ .  $a, b, \sigma$  e  $\gamma$  são parâmetros desconhecidos a ser estimados e  $z_t$  é uma variável i.i.d. com média zero e variância unitária. Este modelo tem, assim, por base a ideia a que já foi feita referência anteriormente sobre a possibilidade do nível da taxa de juro ter efeitos sobre a volatilidade. Neste tipo de formulação, tal como apresentada em (15), a reversão para a média é capturada definindo  $b < 0$  e a leptocurtose ao considerar-se heterocedasticidade condicional, estabelecendo  $\gamma > 0$ . De notar que, neste caso a única informação condicional relevante é o nível das taxas de juro, ou seja,  $\Phi_{t-1} = \{r_{t-1}\}$ .

<sup>14</sup> Devido a Cox, Ingersoll e Ross (1985).

<sup>15</sup> Esta abordagem em tempo discreto serve para aproximar o verdadeiro processo em tempo contínuo.

### 2.2.2. Modelos (G)ARCH lineares

Bollerslev, Chou e Kroner (1992) efectuaram uma análise de algumas aplicações dos modelos do tipo **(G)ARCH** no contexto das taxas de juro. A categoria mais simples desta classe de modelos, **ARCH**, foi introduzida por Engle (1982) com o objectivo de capturar o fenómeno de *volatility clustering*. A ideia base subjacente a este tipo de modelos é que a variância condicional de um choque que ocorre no momento  $t$  é uma função linear dos  $q$  choques passados ao quadrado (ARCH( $q$ )). A forma funcional da variância condicional de  $\varepsilon_t$  assumida nos modelos do tipo ARCH,  $h_t$ , é uma função crescente do choque que ocorreu no período anterior ao quadrado. Assim, se  $\varepsilon_{t-1}$  for elevado (baixo) espera-se que  $h_t$  seja alto (reduzido), e consequentemente,  $\varepsilon_t$  também terá uma variação semelhante<sup>16</sup>. Ou seja, choques de grande (pequena) magnitude, em média, tendem a ser seguidos por choques de grande (pequena) dimensão captando, assim, o comportamento de *volatility clustering* a que já foi feita referência. Tal como demonstrado por Engle, os modelos ARCH conseguem ainda captar a curtose, que é outra característica presente em muitas séries financeiras.

Contudo, este tipo de modelos tem uma limitação relacionada com a sua capacidade para captar a propriedade da persistência. Como a volatilidade tende a exibir uma forte dependência temporal, em aplicações empíricas frequentemente  $q$  tem de ser elevado para conseguir capturar adequadamente a dinâmica da volatilidade condicional. No entanto, nessa situação a estimação dos parâmetros é problemática, dadas as condições de não negatividade e de estacionaridade que precisam de ser impostas, e também porque por vezes, obtêm-se máximos locais e não globais. Engle (1982, 1983) sugeriu um procedimento com vista a reduzir os problemas computacionais associados a esta metodologia, tendo sido tentadas também várias soluções por outros autores. Bollerslev (1986) apresentou ainda outra solução e que tinha igualmente como objectivo a modelação da persistência presente em muitas séries financeiras mas que evita a necessidade de acrescentar muitos termos do resíduo ao quadrado desfasados no modelo. Como alternativa Bollerslev sugeriu simplesmente adicionar à especificação ARCH, termos desfasados da variância condicional, propondo, assim, uma estrutura de parâmetros potencialmente mais parcimoniosa e que designou por **Generalized ARCH**

---

<sup>16</sup> Esta implicação advém da forma como foi definido anteriormente o choque, ou seja,  $\varepsilon_t = z_t \sqrt{h_t}$ .



(GARCH). Tal como referem Bollerslev, Chou e Kroner (1992), na maioria das aplicações o modelo mais simples desta classe, GARCH(1,1), tem-se revelado adequado na modelação da volatilidade. Neste caso concreto,  $h_t$ , assume a seguinte especificação:

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} \quad (16)$$

onde  $\omega$ ,  $\alpha$  e  $\beta$  são parâmetros desconhecidos a estimar<sup>17</sup>. Neste caso os parâmetros devem satisfazer as condições:  $\omega > 0$ ,  $\alpha > 0$  e  $\beta \geq 0$  de forma a garantir que  $h_t \geq 0$ , e  $\alpha$  deve ser estritamente positivo para  $\beta$  ser identificado.

É possível mostrar que este modelo mais parcimonioso corresponde ao ARCH( $\infty$ ), com uma estrutura particular para os parâmetros dos termos  $\varepsilon_t^2$  desfasados.

Nos modelos GARCH, a variância condicional,  $h_t$ , tem uma natureza recursiva e, portanto, contrariamente aos modelos de difusão, a informação condicional relevante é toda a história passada da taxa de juro, ou seja,  $\Phi_{t-1} = \{r_{t-1}, r_{t-2}, r_{t-3}, \dots, r_0\}$ . Contudo,  $h_t | h_{t-1}, \varepsilon_{t-1}$  é independente de toda a história anterior.

O modelo é estacionário em covariância se e sse  $\alpha + \beta < 1$ . Caso em que a variância não condicional de  $\varepsilon_t$  é dada por:

$$\sigma^2 \equiv \text{Var}(\varepsilon_t) = E[\varepsilon_t^2] = \frac{\omega}{1 - \alpha - \beta}. \quad (17)$$

Apesar da variância condicional,  $\text{Var}(\varepsilon_t | \Phi_{t-1})$ , variar no tempo, a variância não condicional,  $\text{Var}(\varepsilon_t)$  é constante. Este tipo de comportamento já havia sido constatado no ponto **2.1.1.** a propósito do comportamento da média, que em termos condicionais também varia no tempo, mas que em termos marginais é constante.

Adicionalmente, é de notar, que quanto mais perto da unidade se encontrar  $\alpha + \beta$ , maior é a persistência, e, portanto, mais duradouro é o efeito dos choques na variância. Bollerslev (1988) derivou a autocorrelação para  $\varepsilon_t^2$ , mostrando que existe um decaimento exponencial para zero, sendo o factor de decaimento  $\alpha + \beta$ . Se este factor

---

<sup>17</sup> O modelo ARCH(1) pode ser visto como um caso particular em que  $\beta = 0$ .

for próximo de um, a autocorrelação decresce apenas gradualmente para zero, traduzindo, assim, a característica de elevada persistência de  $\varepsilon_t^2$ . Este tipo de modelo permite, assim, ultrapassar uma das limitações da formulação ARCH, quando aplicado a séries financeiras, já que consegue captar a característica de persistência.

Existem ainda generalizações deste tipo de modelo, *Generalized ARCH* (GARCH( $p,q$ )). Apesar do interesse teórico destas formulações mais gerais, o modelo mais simples, GARCH(1,1), parece revelar-se suficiente em muitas aplicações (ver Bollerslev, Chou e Kroner (1992)), pelo que não se irá extender análise para modelos mais complexos.

Por vezes quando se aplica o modelo GARCH(1,1) a séries financeiras, as estimativas de  $\alpha$  e de  $\beta$  são tais que a sua soma é próxima ou igual a um. Bollerslev (1986) demonstrou que quando as estimativas dos parâmetros  $\alpha + \beta$  são superiores à unidade, tal implica a não estacionaridade em covariância da variância condicional do processo. Este problema acontece muitas vezes quando se aplica este tipo de modelos a taxas de juro. Nelson (1990) mostrou que nesses casos os choques acumulam-se, no sentido em que  $E[h_{t+m} | h_t] \rightarrow \infty$ , à medida que  $m \rightarrow \infty$ . O que não implica, no entanto, que  $h_{t+m} \rightarrow \infty$  à medida que  $m \rightarrow \infty$ . Frequentemente, contudo, os testes estatísticos tendem a não rejeitar a hipótese nula de que  $\alpha + \beta = 1$ . Quando se verifica esta propriedade ( $\alpha + \beta = 1$ ), e seguindo Engle e Bollerslev (1986), diz-se que  $\varepsilon_t$  segue um processo *Integrated GARCH* (IGARCH). Este tipo de modelo não é estacionário em covariância, contudo, Nelson mostrou que o modelo IGARCH(1,1), poderia, ainda assim, ser estritamente estacionário. Em concreto, foi demonstrado que a condição necessária e suficiente para que  $\varepsilon_t$  seja estritamente estacionário é dada por:

$$E[\log(\beta + \alpha z_t^2)] < 0. \quad (18)$$

Nelson mostrou que esta é uma condição menos exigente que a condição de estacionaridade em covariância, e a que já foi feita referência:  $\alpha + \beta < 1$ . Pode-se provar que:  $\alpha + \beta = 1 \Rightarrow E[\log(\beta + \alpha z_t^2)] < 0$ , de onde se conclui, que o processo do tipo IGARCH é estritamente estacionário. A condição de estacionaridade em covariância

acaba por ser, neste contexto, mais exigente que a de estacionaridade estrita. A primeira exige a existência do momento de segunda ordem, enquanto que, a segunda impõe que toda a estrutura probabilística seja estável ao longo do tempo<sup>18</sup>, independentemente de os momentos serem ou não finitos. Nestas circunstâncias, e contrariamente ao que acontece nos processos integrados na média, que não são estacionários de segunda ordem nem estritamente estacionários, vários estudos mostram que os testes assintóticos continuam válidos.

De notar, que no caso dos modelos IGARCH não é possível definir de forma apropriada a autocorrelação de  $\varepsilon_t^2$ , no entanto Ding e Granger (1996) apresentaram uma forma de aproximar a função de autocorrelação, que à semelhança do verificado no modelo GARCH(1,1), revela um decaimento exponencial para zero.

### 2.2.3. Modelos GARCH não lineares

Existem vários modelos do tipo GARCH não lineares propostos na literatura. Neste trabalho vai ser dado enfoque apenas a um caso particular, que é o enquadramento de MSW com efeitos GARCH (denominado por **MSW-GARCH**) que foi a abordagem seguida por Gray (1996) e cuja metodologia será seguida de perto na aplicação empírica apresentada no **capítulo 3**<sup>19</sup>.

Inicialmente, pensava-se os modelos GARCH com alteração de regime eram intratáveis e impossíveis de estimar devido à dependência da variância condicional de toda a trajectória passada da variável em estudo que existe num modelo deste tipo. A questão é que o modelo **MSW-GARCH** apresenta um problema em termos de estimação que decorre da própria natureza recursiva subjacente ao modelo GARCH. O que acontece é que a distribuição no momento  $t$ , condicional no regime,  $s_t$ , e no conjunto de informação disponível,  $\Phi_{t-1}$ , depende directamente de  $s_t$  e indirectamente de  $s_{t-1}, s_{t-2}, \dots$  devido à dependência temporal inerente aos modelos GARCH com alteração de regime. Tal dependência resulta do facto de a variância condicional no

---

<sup>18</sup> A distribuição conjunta de  $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k)$  é igual à distribuição conjunta de  $(\varepsilon_i, \varepsilon_{i+1}, \dots, \varepsilon_{i+k}), \forall t, k$ ; em particular, as funções densidade de probabilidade são constantes no tempo  $f(\varepsilon_t) = f(\varepsilon_s), \forall t, s$ .

<sup>19</sup> O modelo MSW-GARCH corresponde a assumir a formulação GARCH(1,1) em cada regime, num contexto em que  $s_t$  segue uma cadeia de *Markov* de ordem um. Para uma visão mais completa dos vários tipos de modelos de variância condicional não lineares ver Hentschel (1995), entre outros.

momento  $t$  depender da variância condicional no momento  $t-1$ , que por sua vez é uma função da variância condicional no momento  $t-2$  e, assim, sucessivamente. Consequentemente, a variância condicional no momento  $t$  depende de toda a sequência de regimes até ao momento  $t$ . A função de verosimilhança é construída integrando todas as trajetórias possíveis. Para a  $t$ -ésima observação num modelo com  $k$  regimes, a função de verosimilhança teria  $k^t$  componentes, o que tornaria a estimação intratável em grandes amostras. Em última análise, num modelo com  $k$  regimes, a estrutura que teria de se admitir obrigaria a expandir o número de regimes para  $k^T$ , sendo  $T$  o número total de observações. Para ultrapassar este problema, Cai (1994) e Hamilton e Susmel (1994) desenvolveram um modelo com alteração de regime, mas em que a variância condicional era modelada através de um modelo do tipo ARCH de ordem reduzida.

Posteriormente, Gray (1996) desenvolveu uma metodologia que permitiu resolver o problema da dependência da variância de toda a trajetória passada preservando na sua essência a natureza dos processos GARCH e tornando a sua estimação tratável. Gray introduziu, assim, mais flexibilidade na modelação da variância condicional, aproveitando a capacidade dos modelos GARCH para captar a característica de persistência presente em muitas séries financeiras. Esta aptidão revela-se, particularmente, importante nos casos em que a persistência dos choques individuais varia no tempo. Com efeito, pode acontecer que num determinado período se observem choques de magnitude elevada, mas cujo efeito desaparece rapidamente no tempo, e noutras circunstâncias em que se observe o inverso, ou seja, choques de dimensão mais moderada mas muito persistentes. A construção da função de verosimilhança foi simplificada fazendo uso da natureza recursiva das probabilidades de transição já desenvolvida anteriormente em Gray (1995) e em Hamilton (1994) e que facilitou a estimação de modelos relativamente complicados. A especificação geral do MSW-GARCH e, considerando o caso mais simples de dois regimes, é dada por:

$$h_t = [\omega_1 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}] I[s_t = 1] + [\omega_2 + \alpha_2 \varepsilon_{t-1}^2 + \beta_2 h_{t-1}] I[s_t = 2] \quad (19)$$

onde  $\omega_i$ ,  $\alpha_i$  e  $\beta_i$ , com  $i=1,2$ , são parâmetros a estimar,  $s_t$  representa a cadeia de Markov com probabilidades de transição  $p_{ji}$  tal como definidas no subcapítulo 2.1.2. e  $I[\dots]$  é uma função indicador, igual a um se a condição for verdadeira e igual a zero no

caso oposto. Assim, tal como se vê, é possível distinguir diferentes dinâmicas em termos de volatilidade condicional em função do regime em que se está. Com efeito, num certo regime, o parâmetro ARCH,  $\alpha$ , pode assumir um valor elevado e o parâmetro GARCH,  $\beta$ , um valor pequeno, reflectindo períodos em que se observa uma elevada volatilidade mas com uma persistência relativamente pequena. Por outro lado, também é possível que num outro regime o parâmetro GARCH,  $\beta$ , possa permanecer elevado e o parâmetro ARCH,  $\alpha$ , comparativamente mais reduzido, traduzindo, assim, períodos de elevada persistência mas de menor volatilidade, ou seja, choques com um efeito imediato de menor magnitude.

### **2.3. Modelação da taxa de juro: combinação de modelos**

Os modelos lineares e não lineares abordados no ponto **2.1.** (AR, TAR, SETAR, STAR e MSW), tal como apresentados, normalmente não são apropriados para modelar séries temporais financeiras, como é o caso da taxa de juro, porque todos estes modelos assumem que a variância é constante no tempo. Contudo, vários estudos referem que em muitas séries financeiras acontece precisamente o oposto, sugerindo não só que a variância não é constante ao longo do tempo, como o próprio padrão de heterocedasticidade pode variar consoante o regime. Neste contexto, a generalização desse tipo de modelos de forma a acomodarem a possibilidade de heterocedasticidade surge naturalmente como uma hipótese a tomar em consideração na aplicação empírica.

Com efeito, combinando os modelos apresentados no ponto **2.1.**, para a modelação da média condicional, com os abordados no ponto **2.2.** para a modelação da variância condicional é possível acomodar a hipótese de a variância variar ao longo do tempo nos modelos lineares, bem como ainda a possibilidade de a variância variar em função do regime, no caso específico dos modelos não lineares. Tal como já referido, ao considerar-se heterocedasticidade condicional é possível capturar a eventual leptocurtose presente nos dados. Admitindo o caso mais geral, representado pelo modelo não linear (sendo que o modelo linear pode ser visto como um caso particular deste) e considerando apenas dois regimes (que será abordagem seguida na aplicação empírica) a especificação para variável  $\Delta r_t$  (ou seja,  $r_t - r_{t-1}$ ) seria dada por:

$$\Delta r_t = \begin{cases} \phi_{01} + \phi_{11}r_{t-1} + \dots + \phi_{p1}r_{t-p} + z_t\sqrt{h_{1t}} & \text{se } s_t = 1 \\ \phi_{02} + \phi_{12}r_{t-1} + \dots + \phi_{p2}r_{t-p} + z_t\sqrt{h_{2t}} & \text{se } s_t = 2 \end{cases} \quad (20)$$

onde  $p$  representa o número de defasamentos das variáveis,  $s_t$  é o regime no momento  $t$  (não observável),  $\{z_t\}$  é um ruído branco e  $h_{it}$  com  $i=1,2$ , representa a função de variância condicional. Com efeito, a média condicional encontra-se modelada segundo a formulação AR( $p$ ). Já no que concerne à variância condicional existem várias soluções possíveis, e que a seguir se passam a enumerar:

i) Uma alternativa seria ter em consideração a aparente relação entre a volatilidade e o **nível da taxa de juro**<sup>20</sup> e a que foi feita referência no ponto 2.2.1. no âmbito dos modelos de difusão. A este propósito o modelo unifactorial CIR, que tem implícito  $\gamma=0.5$ , é um dos mais populares. Neste sentido, e considerando a representação em tempo discreto definida por Chan *et al.*, para considerar esta abordagem teria apenas de se definir  $h_{it}$  com a seguinte formulação:

$$h_{it} = \sigma_i^2 r_{t-1}, \text{ com } i=1,2 \quad (21)$$

onde  $i$  representa o regime considerado.

ii) Outra forma possível de captar o facto de a variância variar ao longo do tempo e, simultaneamente, a possibilidade de diferentes regimes assumirem diferentes padrões de heterocedasticidade, seria considerar o modelo **MSW-GARCH**. Tal como já referido, inicialmente pensava-se que este tipo de modelos eram impossíveis de estimar por causa da natureza recursiva subjacente ao modelo GARCH. Contudo, Gray(1996) apresentou uma metodologia (já mencionada no ponto 2.2.3.) que permitiu ultrapassar este problema e a qual vai ser seguida no **capítulo 3**. De acordo com a sugestão de Gray se se assumir a hipótese de normalidade condicional em cada regime<sup>21</sup>, a distribuição condicional da variável  $\Delta r_t$  é, essencialmente, uma mistura de distribuições e que, num contexto de dois regimes, pode ser escrita como:

<sup>20</sup> Esta relação tem sido alvo de uma análise extensiva na literatura, sendo frequentemente referida como “efeito nível”, e refere-se à aparente correlação positiva entre a volatilidade e o nível das taxas de juro, o que significa que em períodos em que as taxas de juro são mais elevadas (baixas), a volatilidade tende a ser maior (menor).

<sup>21</sup> Ou seja, a distribuição condicional de  $\Delta r_t$  no regime 1 é dada por  $N(\mu_{1t}, h_{1t})$  e no regime 2 por  $N(\mu_{2t}, h_{2t})$ , sendo que  $\mu_{it}$  representa a média condicional no regime  $i$  e  $h_{it}$  a variância condicional no regime  $i$ , com  $i = 1, 2$ .

$$\Delta r_t | \Phi_{t-1} \sim \begin{cases} N(\mu_{1t}, h_{1t}) & \text{com probabilidade } p_{1t} \\ N(\mu_{2t}, h_{2t}) & \text{com probabilidade } (1 - p_{1t}) \end{cases} \quad (22)$$

onde  $p_{1t} = \Pr(s_t = 1 | \Phi_{t-1})$ .

Assim, tal como refere Gray, a variância da variação da taxa de juro no momento  $t$  é dada por:

$$\begin{aligned} h_t &= E[\Delta r_t^2 | \Phi_{t-1}] - E[\Delta r_t | \Phi_{t-1}]^2 \\ &= p_{1t}[\mu_{1t}^2 + h_{1t}] + (1 - p_{1t})[\mu_{2t}^2 + h_{2t}] - [p_{1t}\mu_{1t} + (1 - p_{1t})\mu_{2t}]^2. \end{aligned} \quad (23)$$

Como vê, neste enquadramento  $h_t$  não é dependente de toda a história passada<sup>22</sup> e, portanto, pode-se usar a variância condicional desfasada para construir  $h_{1t+1}$  e  $h_{2t+1}$  que seguem processos GARCH. Ou seja, neste caso e tendo por base novamente a especificação da variância condicional, tem-se que:

$$h_{it} = \omega_i + \alpha_i \varepsilon_{it-1}^2 + \beta_i h_{it-1} \text{ com } i=1,2; \quad (24)$$

sendo

$$\begin{aligned} \varepsilon_{it-1} &= \Delta r_{it-1} - E[\Delta r_{it-1} | \Phi_{t-2}] \\ &= \Delta r_{it-1} - [p_{1t-1}\mu_{1t-1} + (1 - p_{1t-1})\mu_{2t-1}] \end{aligned} \quad (25)$$

e

$$\begin{aligned} h_{it-1} &= E[\Delta r_{it-1}^2 | \Phi_{t-2}] - E[\Delta r_{it-1} | \Phi_{t-2}]^2 \\ &= p_{1t-1}[\mu_{1t-1}^2 + h_{1t-1}] + (1 - p_{1t-1})[\mu_{2t-1}^2 + h_{2t-1}] - [p_{1t-1}\mu_{1t-1} + (1 - p_{1t-1})\mu_{2t-1}]^2. \end{aligned} \quad (26)$$

O problema da dependência da variância condicional referido por Cai (1994) e Hamilton e Susmel (1994), surge quando se define  $h_{it}$ , como função de  $\varepsilon_{it-1}$  (erro no período  $t-1$ , associado ao regime  $i$ ) e de  $h_{it-1}$  (variância no período  $t-1$ , no regime  $i$ ). Neste caso, a variância condicional no regime  $i$  no momento  $t$ ,  $h_{it}$ , depende de qual o regime em  $t-1$ . Por sua vez, a variância condicional em  $t-1$  depende do regime em que o processo esteve em  $t-2$ , que, por sua vez, depende do regime em  $t-3$ , e, assim,

<sup>22</sup> De notar que no contexto dos modelos *Markov*, o regime no momento  $t$  não é observável e, portanto,  $\Phi_{t-1}$  não contém  $s_t$  nem valores desfasados de  $s_t$ .

sucessivamente. A especificação sugerida por Gray (1996) e acima apresentada (formulação 22) evita esse problema, pois  $h_{it}$  depende de  $\varepsilon_{t-1}$  e de  $h_{t-1}$ . Basicamente o problema da dependência da variância da sequência de regimes passados, foi resolvido por Gray através de uma única variância condicional (na informação disponível) agregada sobre os regimes.

iii) Adicionalmente, vários autores sugeriram a modelação da variância condicional através da combinação das características dos modelos da classe GARCH com as do modelo de difusão CIR na forma discreta, num quadro de linearidade. Esta abordagem mais geral também foi extendida por Gray (1996) para o contexto não linear usando o mesmo tipo de argumentos já mencionados em ii). Com efeito, Gray sugeriu o modelo **MSW** com efeitos **GARCH** e **CIR**, com dois regimes, mas em que neste caso  $h_{it}$  é representada por:

$$h_{it} = \omega_i + \alpha_i \varepsilon_{t-1}^2 + \beta_i h_{t-1} + \sigma_i^2 r_{t-1} \text{ com } i=1,2 \quad (27)$$

onde  $\omega_i, \alpha_i, \beta_i$  e  $\sigma_i^2$  são parâmetros desconhecidos a ser estimados. Neste caso, a variância condicional é flexível o suficiente para captar certas características, como seja o fenómeno de *volatility clustering*, de persistência dos choques individuais, a par da dependência da volatilidade do nível da taxa de juro.

Na aplicação empírica apresentada no **capítulo 3** irão ser exploradas estas três representações alternativas da variância condicional em combinação com a formulação  $AR(p)$  para a média condicional, na modelação da taxa de juro de longo prazo. O objectivo desta análise é comparar o desempenho de diferentes especificações em termos de *fit in sample* e *out of sample*, em contexto linear e não linear **MSW** (ou seja, em que o indicador de regime latente  $s_t$  segue um processo *Markov* de ordem um).



### 3. Aplicação empírica

Em termos de modelação da média condicional, e tal como já referido, irá considerar-se apenas a formulação  $AR(p)$ , já que este é o tipo de parametrização adoptada pela generalidade dos modelos de taxas de juro<sup>23</sup>. No contexto não linear na média condicional também se assume este tipo de formulação, com a particularidade, de que nesse caso, os parâmetros autoregressivos serem uma função do regime e, portanto, regimes diferentes têm associados diferentes parâmetros e, conseqüentemente, traduzem diferentes dinâmicas ao longo do tempo. Adicionalmente, optou-se ainda por não introduzir variáveis exógenas, como, por exemplo, variáveis macroeconómicas já que o principal objectivo consiste em estudar o desempenho dos modelos não lineares na média condicional por comparação com os modelos lineares quer em termos de ajustamento *in sample* quer de previsão *out of sample*. Como nos modelos do tipo não linear é desejável possuir um elevado número de observações, a inclusão de outras variáveis, como indicadores económicos por exemplo, implicaria considerar uma amostra de menor dimensão, desprezando, assim, uma parte significativa das observações disponíveis para a taxa de juro de longo prazo, cuja série tem início em Janeiro de 1957 e cujo comportamento se pretendia estudar. Tendo em conta que as séries de variáveis macroeconómicas, com frequência mensal<sup>24</sup>, disponíveis para a economia alemã têm um histórico relativamente mais reduzido, a par da exigência neste tipo de modelos de independência das variáveis exógenas face à variável latente indicador de regime,  $s_t$ , justificaram a decisão de incluir nos modelos estudados apenas a variável endógena. De notar que muitos dos modelos de taxas de juro curto prazo, consideram igualmente esta abordagem e que alguns autores argumentam que segundo a teoria da eficiência dos mercados, grande parte da informação disponível já está incorporada nos valores mais recentes da taxa de juro e, por isso, em muitos casos deverá ser suficiente usar os seus valores desfasados.

No que respeita à variância condicional, na literatura existe uma grande variedade de formulações possíveis e, em particular, quando se alarga a análise para o âmbito não linear, são inúmeros os modelos disponíveis. Neste trabalho, o tipo de

---

<sup>23</sup> Em particular a especificação  $AR(1)$  é muito popular já que permite captar a característica de reversão para a média.

<sup>24</sup> A necessidade de dados com frequência mensal decorre do facto de a série em estudo se encontrar nessa base.

modelos estudados restringiram-se basicamente aos analisados por Gray (1996) já que este autor estudou diversas especificações, as quais interpretou como casos particulares de uma formulação mais genérica que designou por *generalized regime-switching* (GRS)<sup>25</sup> e que considerou como tendencialmente mais satisfatória para captar certas características normalmente associadas à volatilidade das taxas de juro. Em conformidade, neste trabalho consideraram-se várias alternativas em termos de modelação da variância condicional (e já enumeradas no ponto 2.3.), começando por modelos relativamente mais simples e terminando com modelos com um grau de complexidade maior. Neste último caso e à semelhança de Gray, incorporou-se efeitos GARCH(1,1) e o nível da taxa de juro, num contexto de probabilidades de transição que variam no tempo.

De notar que quer no caso da média quer da variância condicional, na escolha das parametrizações a incluir na análise foi tida em conta a questão da flexibilidade, mas simultaneamente, a capacidade de manter os modelos parcimoniosos. Com efeito, procurou-se garantir uma adequada especificação dos modelos, mas simultaneamente o cumprimento de alguns requisitos como sejam os modelos serem 1) simples o suficiente para uma fácil estimação, em particular, assegurando uma rápida computação numérica; 2) flexíveis de modo a capturarem diversas situações, nomeadamente, em termos de volatilidade e 3) realísticos, o que neste caso significa não produzirem taxas de juro ou variâncias negativas.

Tal como também já foi mencionado anteriormente, neste trabalho e à semelhança da metodologia seguida por Gray e na generalidade dos modelos de taxas de juro disponíveis na literatura, nos modelos explorados apenas se considera a possibilidade de existirem dois regimes e, portanto,  $s_t$  apenas pode assumir o valor 1 ou 2, consoante se esteja no regime 1 ou 2, respectivamente. Neste contexto, os parâmetros da média e da variância condicional podem assumir dois valores diferentes, dependendo do valor do indicador de regime latente,  $s_t$ . Por sua vez, assume-se que  $s_t$  segue um processo *Markov* de ordem um (ou seja, modelos do tipo **MSW**).

Em termos de procedimento de especificação empírica, no caso dos modelos não lineares na média condicional, Granger (1993) sugere a abordagem do específico para o

---

<sup>25</sup> Esta abordagem incorpora na modelação da variância condicional efeitos GARCH e o nível da taxa de juro, num contexto em que a variável latente,  $s_t$ , segue um processo *Markov* de ordem um com probabilidades de transição que variam no tempo.

geral, sendo este o procedimento adoptado neste trabalho. Por conseguinte, inicialmente começou-se por especificar para a variável em estudo um modelo linear na média condicional apropriado de ordem  $p$ ,  $AR(p)$ . De seguida estimaram-se os parâmetros dos modelos linear e não linear e testou-se a hipótese de linearidade contra a alternativa de não linearidade na média condicional, recorrendo ao teste de Davies. Avaliaram-se os modelos usando os testes de diagnóstico e efectuaram-se as modificações que se consideraram necessárias. Por fim, seleccionaram-se os melhores modelos e procedeu-se à sua avaliação em termos de capacidade descritiva e de previsão *out of sample*.

No próximo subcapítulo efectua-se uma descrição genérica da série em estudo, prosseguindo-se no ponto seguinte para a estimação, inferência e respectivos testes de diagnóstico de diferentes formulações para a modelação da taxa de juro de longo prazo. Nos últimos dois subcapítulos comparam-se os modelos entretanto seleccionados, procurando tirar conclusões sobre qual o mais razoável em termos de ajustamento *in sample* e de previsão *out of sample* para o horizonte de um, dois e três meses.

### 3.1. Base de dados

A variável em estudo corresponde à taxa de juro nominal dos títulos de dívida pública *benchmark* alemães com maturidade de 10 anos, e que no mercados financeiros se designa frequentemente por *yield* do *Bund*<sup>26</sup> a 10 anos. Os dados foram obtidos na Reuters, numa frequência mensal, e correspondem à última taxa de juro (*bid*) cotada pelos contribuidores no final de cada mês. A amostra inclui o período de Janeiro de 1957 até Fevereiro de 2010, englobando, assim, um total de 638 observações. No **gráfico 1** apresenta-se uma ilustração do comportamento da série na amostra considerada e no **gráfico 2** a representação em primeiras diferenças.

---

<sup>26</sup> O *Bund* são obrigações soberanas emitidas pelo Tesouro da Alemanha e dada a sua importância quer em termos de liquidez quer de dimensão constituem um título de referência nos mercados para a área do euro.

Gráfico 1 - Yield do Bund a 10 anos

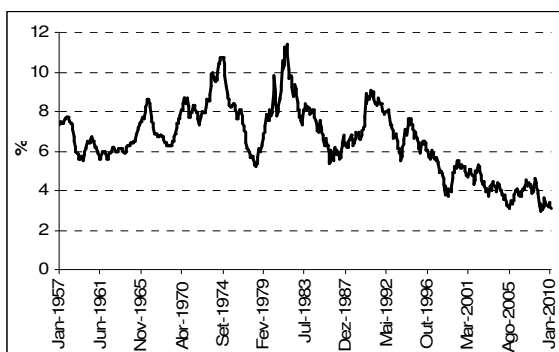
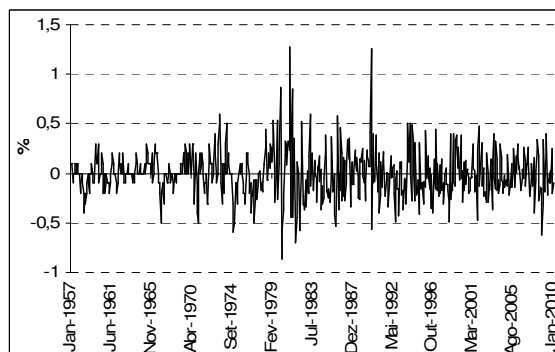


Gráfico 2 – Primeiras diferenças (Bund 10 anos)



Tal como se vê no **gráfico 1** existem momentos na história em que a taxa de juro de longo prazo assumiu valores mais elevados, nomeadamente, entre meados de 1973 e o final de 1974 e entre o final de 1980 e meados de 1982, quando a taxa de juro se fixou acima de 9,0% durante vários meses consecutivos. Por outro lado, mais recentemente, em particular na última década, a taxa de juro tendeu assumir valores mais reduzidos, registando um valor médio em torno de 4%. De notar que desde o início da amostra até ao final dos anos 90, a série evidenciava trajectórias de subida e de descida alternadamente, contudo, desde então parece ter iniciado um movimento de descida prolongado, com pequenas interrupções, que se estendeu até ao final da amostra considerada. Esta tendência recente de taxas de juro mais reduzidas nas maturidades longas nos títulos de dívida pública alemã também se observou noutros soberanos, como é o caso dos EUA, Reino Unido, França, entre outros, e tem sido alvo de estudo por vários autores que procuram identificar os factores que potencialmente poderão explicar este comportamento.<sup>27</sup>

Por sua vez, o **gráfico 2** mostra que em determinados momentos existiram variações significativas da taxa de juro, apresentando um comportamento não regular ao longo do tempo. De notar que no início da amostra e até ao final da década de 70, a volatilidade parece revelar uma maior persistência. Já entre o final de 1979 e o início da década de 80 e nos primeiros meses de 1990 (ano da reunificação alemã) parecem ter existido momentos de elevada volatilidade comparativamente a outros períodos.

Períodos de taxas de juro mais elevadas e de maior volatilidade têm sido identificados por vários autores nas principais economias mundiais. Por exemplo, no caso dos EUA, o período conhecido como *Fed experiment* (1979-1982) e o que engloba o primeiro e o segundo choque petrolífero (1973-1974 e 1979-1980, respectivamente), a

<sup>27</sup> Ver a título de exemplo Ahrend *et al.* (2006).

par do *crash* no mercado bolsista de Outubro de 1987, entre outros, são frequentemente referidos como tal.

Na **tabela 1** apresenta-se um resumo das principais estatísticas descritivas associadas à taxa de juro de longo prazo em primeiras diferenças e de autocorrelação da variável em nível.

**Tabela 1 – Estatísticas descritivas de  $\Delta r_t$  e  $r_t$**

$\Delta r_t$				$r_t$	
Média	-0.007	Curtose	6.311	$\rho_{1^*}$	0.9909
Variância	0.054	Estatística de teste*	17.346	$\rho_{2^*}$	0.9783
Mínimo	-0.870	Assimetria	0.616	$\rho_{3^*}$	0.9648
Máximo	1.280	Estatística de teste**	6.376	$\rho_{12^*}$	0.8169
		<i>Jarque-Bera</i> ***	331.281		
		(pvalue)	0.0000	$\text{corr}(\Delta r_t, r_{t-1})$	-0.045

\* Hipóteses de teste -  $H_0$ : Curtose = 3 vs  $H_1$ : Curtose  $\neq$  3.

\*\* Hipóteses de teste -  $H_0$ : Assimetria = 0 vs  $H_1$ : Assimetria  $\neq$  0.

\*\*\* Hipóteses de teste -  $H_0$ : Curtose = 3 e Assimetria = 0 vs  $H_1$ :  $H_0$  falsa

\* Autocorrelação de ordem 1, 2, 3 e 12, respectivamente.

Em termos médios, a variável em primeiras diferenças tende apresentar um valor próximo de zero, e há indícios de excesso de curtose e de enviesamento positivo face à distribuição normal, o que revela que a sua distribuição deverá ser leptocúrtica e com enviesamento à direita. Estes factos são confirmados pelo teste *Jarque-Bera*, cujos resultados apontam para a não normalidade uma vez que é rejeitada a hipótese nula de teste. Adicionalmente é de notar a elevada persistência da variável em nível, tal como sugere o valor do coeficiente de autocorrelação para os vários desfasamentos.

### 3.2. Estimação, inferência e testes de diagnóstico

No caso dos modelos lineares na média condicional apenas é necessários estimar os parâmetros da média e variância condicional como é usual. Já no caso dos modelos não lineares, como o processo *Markov*,  $s_t$ , não é observado, têm de ser estimados não só os parâmetros da média e variância condicional associados a cada regime, como também as probabilidades de transição entre regimes, a par da estimativa da probabilidade de cada estado ocorrer em cada momento de tempo da amostra. Tal como já referido anteriormente, o problema de estimação dos parâmetros será discutido

apenas no âmbito de modelos **MSW** com dois regimes e, assumindo o mesmo número de parâmetros AR em cada regime.

Neste contexto, relembando novamente a especificação do modelo **MSW** (fórmula 20 apresentada no ponto 2.3.) tem-se:

$$\Delta r_t = \begin{cases} \phi_{01} + \phi_{11}r_{t-1} + \dots + \phi_{p1}r_{t-p} + z_t\sqrt{h_{1t}} & \text{se } s_t = 1 \\ \phi_{02} + \phi_{12}r_{t-1} + \dots + \phi_{p2}r_{t-p} + z_t\sqrt{h_{2t}} & \text{se } s_t = 2 \end{cases}. \quad (28)$$

Quando  $\Delta r_t$  está sujeita a alterações entre dois regimes, a sua função densidade de probabilidade condicional na informação disponível é dada por:

$$\begin{aligned} f(\Delta r_t | \Phi_{t-1}) &= f(\Delta r_t | s_t = 1, \Phi_{t-1})p_{1t} + f(\Delta r_t | s_t = 2, \Phi_{t-1})p_{2t} \\ &= g_{1t}p_{1t} + g_{2t}p_{2t} \\ &= g_{1t}p_{1t} + g_{2t}(1 - p_{1t}) \end{aligned} \quad (29)$$

onde  $p_{1t} = \Pr(s_t = 1 | \Phi_{t-1})$ ,  $p_{2t} = \Pr(s_t = 2 | \Phi_{t-1}) = 1 - p_{1t}$ . Ou seja, a função densidade de probabilidade de  $\Delta r_t$  condicional na informação disponível corresponde à média ponderada das funções densidade de probabilidade condicionadas aos respectivos regimes e informação disponível, sendo os ponderadores dados pela probabilidade de se estar nesse regime num certo momento de tempo condicional na informação disponível, ou seja,  $p_{1t}$  e  $p_{2t}$ . Assim, quanto maior a probabilidade de se estar num certo regime num determinado momento de tempo,  $\Pr(s_t = i | \Phi_{t-1})$ , maior a dependência da função  $f(\Delta r_t | \Phi_{t-1})$  da função densidade probabilidade condicionada associada a esse regime,  $f(\Delta r_t | s_t = i, \Phi_{t-1})$ .

O objectivo é calcular  $f(\Delta r_t | \Phi_{t-1})$ , mas tal como se vê pela especificação (29) para tal é preciso obter  $g_{1t}, g_{2t}$  e  $p_{1t}$ . Para facilitar a compreensão do procedimento em causa irá descrever-se em **i)** como obter  $g_{it}$  e em **ii)** como calcular  $p_{1t}$ , e assim, ter a formulação  $f(\Delta r_t | \Phi_{t-1})$  completa. Com efeito:

**i)** se se assumir normalidade condicional em cada regime (ver Gray (1996) para mais detalhes), ou seja,  $z_t \sim N(0,1)$ , a função densidade de probabilidade de  $\Delta r_t$  condicionada na informação disponível e no regime pode ser escrita como:

$$g_{it} = f(\Delta r_t | s_t = i, \Phi_{t-1}) = \frac{1}{\sqrt{2\pi h_{it}}} \exp \left\{ \frac{-(\Delta r_t - \mu_{it})^2}{2h_{it}} \right\} \quad (30)$$

sendo que  $\mu_{it} = \phi_{0i} + \phi_{1i}r_{t-1} + \dots + \phi_{pi}r_{t-p}$  com  $i=1,2$  e  $h_{it}$  representa a variância condicional. No caso dos modelos abordados na aplicação empírica,  $h_{it}$  poderá ser uma função do nível da taxa de juro, ou assumir a representação GARCH(1,1) ou alternativamente ser uma combinação das duas (efeito GARCH e nível).

ii) relativamente a  $p_{1t}$  e seguindo a abordagem de Hamilton (1988, 1989, 1990), a variável latente  $s_t$ , pode ser parametrizada como um processo *Markov* de primeira ordem, o que implica que o regime num determinado momento,  $s_t$ , depende do regime do processo no momento anterior,  $s_{t-1}$ . Com efeito, num contexto de dois regimes e recorrendo à regra da probabilidade total, têm-se que:

$$p_{1t} = \Pr(s_t = 1 | \Phi_{t-1}) = \Pr(s_t = 1 | s_{t-1} = 1, \Phi_{t-1}) \Pr(s_{t-1} = 1 | \Phi_{t-1}) + \Pr(s_t = 1 | s_{t-1} = 2, \Phi_{t-1}) \Pr(s_{t-1} = 2 | \Phi_{t-1}). \quad (31)$$

Para completar a definição de  $p_{1t}$  é necessário, ainda, caracterizar as probabilidades de transição e a este respeito colocaram-se duas hipóteses. A abordagem mais simples insere-se no enquadramento de Hamilton (que é a mais comum na literatura) e assume que  $s_t$  segue uma cadeia de *Markov* homogénea, ou seja, as **probabilidades de transição de Markov são constantes**. Alternativamente também se pode assumir outra abordagem, e que também foi usada por Gray (1996), que tem implícito que as **probabilidades de transição de Markov variam no tempo**.

No contexto de **probabilidades de transição constantes**,  $s_t$  dado  $s_{t-1}$ , não depende de  $\Phi_{t-1}$ , ou seja,  $\Pr(s_t = i | s_{t-1} = j, \Phi_{t-1}) = \Pr(s_t = i | s_{t-1} = j) = p_{ji}$ , o que se traduz numa matriz de probabilidades de transição **constantes no tempo**,  $M$ , que pode ser definida como:

$$M = \begin{pmatrix} P & (1-P) \\ (1-Q) & Q \end{pmatrix}, \quad (32)$$

$$\begin{aligned}
P &= \Pr[s_t = 1 | s_{t-1} = 1] = p_{11}, \\
(1-P) &= \Pr[s_t = 2 | s_{t-1} = 1] = p_{12}, \\
Q &= \Pr[s_t = 2 | s_{t-1} = 2] = p_{22}, \\
(1-Q) &= \Pr[s_t = 1 | s_{t-1} = 2] = p_{21}.
\end{aligned}
\tag{33}$$

De notar que a matriz  $M$ , não depende de  $t$ , pois o que é relevante no cálculo das probabilidades não é o momento em que são calculadas, mas sim o desfasamento temporal entre as variáveis (neste caso um desfasamento). Neste enquadramento, a expressão (31) pode ser reescrita como:

$$\begin{aligned}
p_{1t} &= \Pr(s_t = 1 | \Phi_{t-1}) = P \times \Pr(s_{t-1} = 1 | \Phi_{t-1}) + (1-Q) \times \Pr(s_{t-1} = 2 | \Phi_{t-1}) \\
&= P \times \Pr(s_{t-1} = 1 | \Phi_{t-1}) + (1-Q) \times (1 - \Pr(s_{t-1} = 1 | \Phi_{t-1})).
\end{aligned}
\tag{34}$$

Por sua vez, a expressão (34) exige o cálculo de  $\Pr(s_{t-1} = 1 | \Phi_{t-1})$ . Assim, tem-se que:

$$\begin{aligned}
\Pr(s_{t-1} = 1 | \Phi_{t-1}) &= \Pr(s_{t-1} = 1 | \Delta r_{t-1}, \Phi_{t-2}) \quad \rightarrow \quad \text{por definição} \\
&= \frac{j(\Delta r_{t-1}, s_{t-1} = 1 | \Phi_{t-2})}{f(\Delta r_{t-1} | \Phi_{t-2})} \quad \rightarrow \quad \text{usando a regra } \Pr(A|B,C) = \Pr(A,B|C)/\Pr(B|C) \\
&= \frac{f(\Delta r_{t-1} | s_{t-1} = 1, \Phi_{t-2}) \Pr(s_{t-1} = 1 | \Phi_{t-2})}{\sum_{i=1}^2 f(\Delta r_{t-1} | s_{t-1} = i, \Phi_{t-2}) \Pr(s_{t-1} = i | \Phi_{t-2})} \quad \rightarrow \quad \frac{\text{usando a regra } \Pr(A,B|C) = \Pr(A|B,C) \times \Pr(B|C)}{\text{por analogia à expressão (29)}} \\
&= \frac{g_{1t-1} p_{1t-1}}{\sum_{i=1}^2 g_{it-1} p_{it-1}}
\end{aligned}
\tag{35}$$

sendo que  $j$  é a função de probabilidade conjunta de  $(\Delta r_{t-1}, s_{t-1})$ .

Com efeito, a expressão (34), após simplificação pode ser reescrita como:

$$p_{1t} = P \times \left( \frac{g_{1t-1} p_{1t-1}}{\sum_{i=1}^2 g_{it-1} p_{it-1}} \right) + (1-Q) \times \left( \frac{g_{2t-1} (1 - p_{1t-1})}{\sum_{i=1}^2 g_{it-1} p_{it-1}} \right).
\tag{36}$$

Com a definição de **i)** e **ii)** a expressão (29) ficou completamente definida. Neste sentido,  $p_{1t}$ , pode ser interpretado como um parâmetro não observado da distribuição condicional e que envolve uma estrutura recursiva que a seguir se expõe. Dadas as condições iniciais, ou seja, os respectivos valores no momento zero,  $h_{10}$ ,  $h_{20}$ ,  $p_{10}$ ,  $\epsilon_0^2$  e os valores iniciais arbitrários definidos para os parâmetros a estimar, é possível obter



$g_{10}$  e  $g_{20}$  (através da expressão (30)) e, posteriormente,  $p_{11}$  (tal como se vê na expressão (36)). Este processo recursivo repete-se para  $t = 2$ , calculando  $p_{12}$  a partir dos valores  $g_{11}$ ,  $g_{21}$  e  $p_{11}$  e, assim, sucessivamente até se obter  $p_{1t}$ .

No caso alternativo de **probabilidades de transição Markov que variam no tempo**, e contrariamente ao caso anterior,  $s_t$  dado  $s_{t-1}$  depende de  $\Phi_{t-1}$ . Os regimes também seguem um processo de Markov de primeira ordem, por isso, a estrutura recursiva de  $p_{1t}$  é semelhante à que foi apresentada anteriormente. A particularidade é que neste caso,  $P$  e  $Q$  variam no tempo, ou seja, passa-se a ter em substituição de  $P$  e  $Q$  na expressão (34),  $P_t$  e  $Q_t$ . Neste contexto tem de se assumir uma forma funcional para  $P_t$  e  $Q_t$ , que podem ser arbitrárias, desde que se garanta que  $P_t, Q_t \in (0,1)$  em qualquer momento de tempo. No caso de dois regimes esta condição é garantida se se modelarem estes dois parâmetros com base na representação logística ou *probit*. Seguindo a abordagem de Gray (1996), vai-se assumir a formulação *probit*, admitindo que as probabilidades de transição dependem do nível da taxa de juro de longo prazo no momento anterior,  $r_{t-1}$ . Por conseguinte, tem-se que,  $P_t = \Phi(c_1 + d_1 r_{t-1})$  e  $Q_t = \Phi(c_2 + d_2 r_{t-1})$ , onde  $\Phi(\cdot)$  representa a função de distribuição normal cumulativa,  $c_i$  e  $d_i$ ,  $i = 1,2$ , são parâmetros desconhecidos a estimar.

Após a especificação da média condicional,  $\mu_{it}$ , da variância condicional,  $h_{it}$ , a par da dinâmica de transição entre regimes, a **função do logaritmo de verosimilhança**, segundo o enquadramento de Gray, pode ser escrita como:

$$L = \sum_{t=1}^T \log \left( \sum_{i=1}^2 p_{it} g_{it} \right) = \sum_{t=1}^T \log (p_{1t} g_{1t} + (1 - p_{1t}) g_{2t}). \quad (37)$$

Definindo  $\theta$  como o vector de parâmetros a estimar, o estimador de máxima verosimilhança é dado por:

$$\hat{\theta}_T = \arg \max_{\theta} \sum_{t=1}^T l_t(\theta) \quad (38)$$

onde  $l_t(\theta) = \log f(\Delta r_t | \Phi_{t-1}; \theta) = \log (p_{1t} g_{1t} + (1 - p_{1t}) g_{2t})$ .

De notar que o vector com os  $T$  valores da função do logaritmo de verosimilhança, à semelhança da variância nos processos GARCH e do parâmetro  $p_{1t}$ , também é construído recursivamente. Assim, à medida que se constrói o vector da variância,  $h_{1t}$  e  $h_{2t}$ , e o vector  $p_{1t}$  para cada observação, é possível calcular sucessivamente o vector função do logaritmo de verosimilhança, até se obter  $f(\Delta r_T | \Phi_{T-1})$  (ver expressão (30), (36) e (37)).

Em termos de **inferência**, Gray (1995) demonstrou a consistência e normalidade assintótica dos estimadores de quase máxima verosimilhança para alguns modelos com alteração de regime, sob condições de regularidade relativamente modestas. Com efeito, os testes de significância individuais para os parâmetros da média e da variância condicional, e à semelhança do procedimento seguido por Gray (1996), foram feitos como habitualmente, ou seja, usando a estatística  $t$  e os respectivos valores críticos assintóticos. O teste do rácio de verosimilhança (RV) foi usado para comparar diferentes formulações dos modelos lineares entre si e também modelos não lineares. Contudo, nem sempre foi possível usar os testes assintóticos habituais<sup>28</sup>, nomeadamente, quando se pretendia testar a significância do segundo regime face ao modelo com apenas um regime. A título exemplificativo, tendo por referência a especificação (28), quando as hipóteses nula e alternativa sob teste são dadas por:  $H_0: \phi_{01} = \phi_{02}, \phi_{11} = \phi_{12}, \dots, \phi_{p1} = \phi_{p2}, h_1 = h_2$  e  $H_1: H_0$  é falsa, não se podem aplicar os testes assintóticos habituais. Esta impossibilidade resulta do facto de sob a hipótese nula de existência de um único regime, os parâmetros associados ao segundo regime não serem identificados<sup>29</sup>. Neste caso, a estatística RV deixa de ter distribuição  $\chi^2$  sob  $H_0$ . Uma solução possível para lidar com este problema sugerida na literatura, é o teste proposto por Davies (1987). Este teste obtém um limite superior para o nível de significância do teste do rácio de verosimilhança quando são identificáveis apenas  $q$  parâmetros sob a hipótese alternativa. No caso de se testar a hipótese alternativa de dois regimes, estes parâmetros são precisamente  $P$  e  $Q$ . Com efeito, o teste tem como formulação:

<sup>28</sup> Como sejam, o rácio de verosimilhança, o teste de *Wald* ou o multiplicador de *Lagrange*.

<sup>29</sup> O *score* com respeito a esses parâmetros é identicamente nulo e, por conseguinte, a matriz de informação de Fisher resulta singular.

$$\Pr(\sup LR > LR_{obs}) \leq \Pr(\chi_q^2 > LR_{obs}) + V \times LR_{obs}^{(q-1)/2} \exp\{-LR_{obs}\} \times \frac{2-q/2}{\Gamma(q/2)} \quad (39)$$

sendo que  $LR_{obs} = 2(\log L_n^1 - \log L_n^0)$  e  $V = 2\sqrt{LR_{obs}}$ .  $\Gamma$  representa a função gama e  $L_n^1$  e  $L_n^0$  são os valores do logaritmo da função da verosimilhança sob  $H_1$  e  $H_0$ , respectivamente. A implementação do teste exige, assim, a estimação do modelo com um e dois regimes. De notar que, este teste não fornece indicações sobre o verdadeiro *p-value*, mas sim apenas um limite superior para esse valor.

Em relação aos **testes de diagnóstico**, considerou-se o teste de autocorrelação de *Ljung-Box*<sup>30</sup> aplicado aos resíduos estandardizados e aos resíduos estandardizados ao quadrado com o objectivo de indirectamente verificar se o modelo está ou não correctamente especificado. Com efeito, existe evidência a favor de que a média e a variância condicional estão bem modeladas, se o teste de autocorrelação dos resíduos estandardizados e dos resíduos estandardizados ao quadrado, respectivamente, apontarem para a não rejeição da hipótese nula. Adicionalmente, considerou-se o teste de *Jarque-Bera*<sup>31</sup> com o intuito de avaliar a normalidade dos resíduos estandardizados. De acordo com este último teste, a não rejeição da hipótese nula revela a presença de indícios de normalidade.

De seguida apresenta-se os resultados de estimação de vários modelos para a taxa de juro de longo prazo, procedendo-se a uma breve análise e interpretação de cada tipo de modelo. O objectivo é perceber qual a especificação que permite captar melhor as características da média e da variância condicional da taxa de juro de longo prazo. De notar que os modelos apresentados foram todos estimados numa primeira fase

---

<sup>30</sup> O teste de *Ljung-Box* com  $k$  desfasamentos tem como hipóteses de teste:  $H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0$  vs

$H_1 : \exists \rho_j \neq 0 \quad j = 1, 2, \dots, k$ . A estatística de teste,  $Q_{LB}$ , é dada por:  $Q_{LB} = T(T+2) \sum_{j=1}^k \frac{\hat{\tau}_j^2}{T-j}$ , onde  $\tau_j$  representa a autocorrelação de ordem  $j$  e  $T$  o número de observações. Sob a hipótese nula,  $Q_{LB}$  tem, assintoticamente, uma distribuição  $\chi^2$  com  $k$  graus de liberdade.

<sup>31</sup> O teste de *Jarque-Bera* assume como  $H_0$  assimetria nula e curtose igual a três e sob  $H_1$  que  $H_0$  é falsa. A estatística de teste,  $JB$ , é dada por:  $JB = T \left( \frac{(\hat{k}-3)^2}{24} + \frac{sk^2}{6} \right) \sim \chi_{(2)}^2$  sob  $H_0$ , onde  $T$  é o número de observações,  $k$  a curtose e  $sk$  a assimetria.

assumindo como especificação para a média condicional, a formulação AR(1), contudo, todos falharam os testes de diagnóstico de especificação da média condicional (ou seja, o teste de *Ljung-Box* aplicado aos resíduos estandardizados), existindo indícios de má modelação para os níveis de significância usuais (1, 5 e 10%). De seguida reestimaram-se os modelos, mas considerando a formulação AR(2), tendo-se verificado evidência a favor da não rejeição da hipótese nula e, portanto, de que a média condicional se encontra bem especificada. Por este motivo, apenas serão apresentados os resultados dos modelos que assumem a formulação AR(2) em termos de média condicional.

Todos os modelos foram estimados por quase-máxima verosimilhança, usando o módulo CML (*Constrained Maximum Likelihood*) do GAUSS<sup>32</sup>. No que se refere ao método de optimização usado na estimação dos parâmetros, o programa vai iterando entre diferentes algoritmos<sup>33</sup>, na tentativa de encontrar o que melhor permite maximizar o logaritmo da função de verosimilhança.

### 3.2.1. Modelo com o nível da taxa de juro<sup>34</sup>

No âmbito desta classe de modelos será estimado o modelo linear na média condicional, em que se admite a existência de um único regime, e o não linear na média condicional, em que se assume a presença de dois regimes. Por uma questão de simplificação da exposição, serão designados doravante por **modelos da classe 1**, sendo o modelo linear denominado por **1.a)** e o modelo que admite alterações de regime do tipo *Markov* de primeira ordem, por **1.b)**. Em ambos os casos admite-se a possibilidade de heterocedasticidade condicional, com a dinâmica da volatilidade a ser representada

---

<sup>32</sup> O desvio padrão das estimativas dos parâmetros reportado nas tabelas é calculado a partir da diagonal da matriz de variância-covariância do estimador (*heteroskedastic-consistent covariance matrix*). Para mais detalhes ver White (1980).

<sup>33</sup> Por defeito, o processo de iteração é iniciado com o algoritmo de Newton-Raphson (NEWTON), mas durante a sequência de iterações outros algoritmos, como os desenvolvidos por Broyden, Fletcher, Goldfarb e Shanno (BFGS); Davidon, Fletcher e Powell (DFP) ou Berndt, Hall, Hall e Hausman (BHHH) podem integrar esse processo.

<sup>34</sup> Seguindo abordagem de Chan *et al* (1992), neste subcapítulo apresenta-se a estimação da equação diferencial estocástica discretizada tal como já apresentada no ponto 2.2.1.. Contudo, e seguindo a metodologia de Gray (1996), o parâmetro  $\gamma$  não será considerado como um parâmetro livre a estimar, mas sim restrito à partida a um valor fixo. Tendo por referência a formulação CIR para as taxas de juro de curto prazo, Gray optou por definir no seu modelo  $\gamma = 0.5$ . Neste trabalho inicialmente foram tidos em consideração três alternativas possíveis para este parâmetro:  $\gamma = 0.25, 0.5$  e  $0.75$ . Nas experiências realizadas, os resultados de estimação dos diferentes modelos não revelaram diferenças significativas entre si. Tendo em conta este facto e dada a popularidade dos modelos CIR no contexto de modelação das taxas de juro optou-se por seguir a mesma abordagem de Gray fixando  $\gamma$  em 0.5, facilitando simultaneamente a interpretação dos resultados.

como uma função do nível das taxas de juro de longo prazo. Este tipo de formulação é a mais simples que foi considerada neste trabalho, já que os modelos apresentados nos próximos pontos tendem a apresentar um grau crescente de complexidade, quer pelo número mais elevado de parâmetros a estimar quer pelas dinâmicas mais sofisticadas de volatilidade que tentam captar.

Na **tabela 2**<sup>35</sup> reportam-se os resultados de estimação destes dois modelos<sup>36</sup>.

---

<sup>35</sup> As raízes do polinómio AR reportadas nas tabelas ao longo de todo o **capítulo 3** foram calculadas no Gauss usando o comando *polyroot*. O objectivo foi aferir sobre a estacionaridade do processo, situação em que as raízes da equação característica se deverão situar fora do círculo unitário. Com efeito, como nos modelos sob análise, a média condicional foi modelada com a formulação do tipo AR(2), as soluções 1 e 2 dizem respeito às raízes ( $L$ ) da equação característica de um polinómio de ordem dois. Assim, no caso dos modelos que consideram apenas um regime, o polinómio é dado por:  $\phi_2(L) = (1 - (a_1 + 1)L - a_2L^2)$  e reportam-se as soluções da expressão:  $(1 - (a_1 + 1)L - a_2L^2) = 0$ . Por sua vez, no caso dos modelos com dois regimes, tem-se um polinómio de ordem dois associado a cada regime, pelo que os valores reportados dizem respeito às soluções da expressão:  $(1 - (a_{11} + 1)L - a_{21}L^2) = 0$  no regime 1 e da equação:  $(1 - (a_{12} + 1)L - a_{22}L^2) = 0$  no caso do regime 2.

<sup>36</sup> Em anexo apresenta-se informação adicional associada aos modelos estimados, nomeadamente, a representação gráfica da média e da variância condicional, FAC e FACP dos resíduos estandardizados e *smoothed probability*.

**Tabela 2 – Resultados de estimação do modelo 1.a) e 1.b)**

**Especificação do modelo**

Modelo 1.a)

$$\Delta r_t = a_0 + a_1 r_{t-1} + a_2 r_{t-2} + \sigma_1 \sqrt{r_{t-1}^{2.05}} z_t$$

$$z_t \sim i.i.d. N(0,1)$$

$$\Delta r_t | \Phi_{t-1} \sim N(a_0 + a_1 r_{t-1} + a_2 r_{t-2}, \sigma_1^2 r_{t-1})$$

**Resultados da estimação**

Mean log-likelihood				
0.0844				
	estimativas	desvio-padrão	estatística t	p-value
$a_0$	0.0418	0.0374	1.1164	0.2642
$a_1$	0.1808	0.0570	3.1695	0.0015
$a_2$	-0.1880	0.0560	-3.3584	0.0008
$\sigma_1$	-0.0884	0.0035	-25.4875	0.0000

**Raízes do polinómio AR**

solução 1	solução 2
5.2719	1.0090

**Critérios de informação**

AIC	-0.1562
BIC	-0.1282

**Testes de diagnóstico**

Teste de autocorrelação de *Ljung-Box*

	Resíduos <i>standardizados</i>		Resíduos <i>standardizados</i> ao quadrado	
	Estatística	p-value	Estatística	p-value
<i>Lag1</i>	0.0070	0.9335	51.0286	0.0000
<i>Lag2</i>	0.2021	0.9039	66.0243	0.0000
<i>Lag3</i>	2.4548	0.4835	76.2397	0.0000
<i>Lag4</i>	4.9834	0.2890	81.2212	0.0000
<i>Lag5</i>	5.1566	0.3971	84.0378	0.0000
<i>Lag6</i>	5.5282	0.4781	92.5693	0.0000
<i>Lag12</i>	11.8086	0.4612	129.9623	0.0000

Estatísticas e teste de normalidade dos resíduos *standardizados*

Curtose	4.9001
Assimetria	0.4344
<i>Jarque-Bera</i>	115.6724
(p value)	0.0000

**Especificação do modelo**

Modelo 1.b) Probabilidades de transição constantes

$$\Delta r_t = \begin{cases} a_{01} + a_{11} r_{t-1} + a_{21} r_{t-2} + \sigma_1 \sqrt{r_{t-1}^{2.05}} z_t & \text{se } s_t = 1 \\ a_{02} + a_{12} r_{t-1} + a_{22} r_{t-2} + \sigma_2 \sqrt{r_{t-1}^{2.05}} z_t & \text{se } s_t = 2 \end{cases}$$

$$z_t \sim i.i.d. N(0,1)$$

$$\Delta r_t | \Phi_{t-1} \sim \begin{cases} N(a_{01} + a_{11} r_{t-1} + a_{21} r_{t-2}, \sigma_1^2 r_{t-1}) & \text{com } p_{1t} \\ N(a_{02} + a_{12} r_{t-1} + a_{22} r_{t-2}, \sigma_2^2 r_{t-1}) & \text{com } 1-p_{1t} \end{cases}$$

**Resultados da estimação**

Mean log-likelihood				
0.1896				
	estimativas	desvio-padrão	estatística t	p-value
$a_{01}$	0.1007	0.0581	1.7348	0.0828
$a_{02}$	0.0366	0.0449	0.8159	0.4145
$a_{11}$	0.5363	0.0650	8.2556	0.0000
$a_{12}$	0.0872	0.0685	1.2740	0.2027
$a_{21}$	-0.5505	0.0661	-8.3336	0.0000
$a_{22}$	-0.0945	0.0662	-1.4285	0.1531
$\sigma_1$	0.0529	0.0029	18.5187	0.0000
$\sigma_2$	0.1042	0.0047	22.2091	0.0000
P	0.9978	0.0031	323.1513	0.0000
Q	0.9987	0.0019	531.7596	0.0000

**Raízes do polinómio AR**

Regime 1	solução 1	solução 2
Regime 1	1.7565	1.0341
Regime 2	10.4917	1.0082

**Critérios de informação**

AIC	-0.3478
BIC	-0.2777

**Testes de diagnóstico**

Teste de autocorrelação de *Ljung-Box*

	Resíduos <i>standardizados</i>		Resíduos <i>standardizados</i> ao quadrado	
	Estatística	p-value	Estatística	p-value
<i>Lag1</i>	0.0512	0.8210	15.8525	0.0001
<i>Lag2</i>	2.1671	0.3384	18.1399	0.0001
<i>Lag3</i>	2.4347	0.4872	20.2239	0.0002
<i>Lag4</i>	5.2270	0.2648	24.4878	0.0001
<i>Lag5</i>	5.2350	0.3879	28.5788	0.0000
<i>Lag6</i>	5.3032	0.5056	34.6947	0.0000
<i>Lag12</i>	8.2488	0.7654	50.8643	0.0000

Estatísticas e teste de normalidade dos resíduos *standardizados*

Curtose	3.9162
Assimetria	0.3903
<i>Jarque-Bera</i>	38.3947
(p value)	0.0000

No modelo com apenas um regime, **1.a)** e no que se refere aos parâmetros da média condicional, parecem existir indícios de significância estatística das estimativas de todos os parâmetros, para qualquer um dos níveis de significância habituais<sup>37</sup>, com a exceção da constante,  $a_0$ . A taxa de juro média de longo prazo implícita é de 5.8%<sup>38</sup>.

<sup>37</sup> Os níveis de significância usuais são 1, 5 e 10%.

<sup>38</sup> A média de longo prazo pode ser obtida como:  $-a_0 / (a_1 + a_2)$ .

No caso do modelo **1.b)**, ou seja, com dois regimes, as estimativas dos parâmetros da média condicional do regime 1 são estatisticamente significantes (considerando  $\alpha = 10\%$ ), enquanto que as do regime 2 parecem ser estatisticamente insignificantes. A taxa de juro média de longo prazo implícita no regime 1 é de 7.1%, enquanto que no regime 2 é de 5.0%. Adicionalmente, é de notar que no regime 1, a estimativa do parâmetro  $\sigma$  é cerca de metade da estimativa do regime 2, sendo estatisticamente significativas em ambos os casos, o que parece indiciar comportamentos distintos dos dois regimes em termos de volatilidade. A estimativa de P e Q é superior a 0.95, apontando para a existência de regimes muito persistentes.

Os testes de diagnóstico parecem indiciar que quer no modelo **1.a)** quer no **1.b)**, a média condicional está bem especificada mas o mesmo não se pode concluir quanto à modelação da variância condicional, já que os testes de *Ljung-Box* sugerem a presença de autocorrelação nos resíduos estandardizados ao quadrado. Adicionalmente, o teste de *Jarque-Bera* indicia que a distribuição dos resíduos estandardizados deverá caracterizar-se por uma assimetria positiva e caudas mais pesadas do que a distribuição normal. Apesar da rejeição da hipótese de normalidade em ambos os casos, no modelo com dois regimes, **1.b)**, os desvios da normalidade aparentemente são menos pronunciados, tal como sugere o decréscimo significativo da estatística de teste *Jarque-Bera*.

Globalmente, comparando os dois modelos, **1.a)** e **1.b)**, ambos parecem ter uma adequada especificação da média condicional, contudo, falham na modelação da volatilidade condicional. O modelo com dois regimes, **1.b)**, parece produzir um melhor ajustamento aos dados comparativamente ao modelo com apenas um regime, **1.a)**, já que os critérios AIC e BIC diminuíram de forma significativa, e o valor médio do logaritmo da verosimilhança, subiu de 0.0844, para 0.1896. De acordo com o teste de Davies, a hipótese nula é rejeitada para qualquer nível de significância. Neste contexto, parecem existir indícios que apontam para a superioridade do modelo com dois regimes face ao modelo com apenas um regime.

No próximo subtópico apresenta-se outro tipo de especificação, assumindo-se agora uma parametrização para a variância condicional que admite a presença de efeitos GARCH.

### 3.2.2. Modelo com efeitos GARCH

Os modelos que se inserem neste enquadramento serão referidos no remanescente deste trabalho como pertencentes à classe **2**, sendo que o modelo que assume apenas um regime (linear na média condicional), será designado por **2.a**) e, por seu turno, o modelo que admite a possibilidade de dois regimes por **2.b**). Em ambos os casos e à semelhança dos modelos da classe **1** admite-se a possibilidade de heterocedasticidade condicional, contudo, neste caso assume-se que esta é uma função de efeitos GARCH(1,1).

Na **tabela 3** reportam-se os resultados de estimação destes dois modelos.



**Tabela 3 – Resultados de estimação do modelo 2.a) e 2.b)**

**Especificação do modelo**

Modelo 2.a)

$$\Delta r_t = a_0 + a_1 r_{t-1} + a_2 r_{t-2} + (\sqrt{b_0 + b_1 \varepsilon_{t-1}^2 + b_2 h_{t-1}}) z_t$$

$$z_t \sim i.i.d. N(0,1)$$

$$\Delta r_t | \Phi_{t-1} \sim N(a_0 + a_1 r_{t-1} + a_2 r_{t-2}, h_t)$$

$$h_t = b_0 + b_1 \varepsilon_{t-1}^2 + b_2 h_{t-1}$$

**Resultados da estimação**

Mean log-likelihood 0.1690				
	estimativas	desvio-padrão	estatística t	p-value
$a_0$	0.0723	0.0329	2.1960	0.0281
$a_1$	0.3095	0.0447	6.9175	0.0000
$a_2$	-0.3218	0.0447	-7.2063	0.0000
$b_0$	0.0013	0.0009	1.4561	0.1454
$b_1$	0.1706	0.0531	3.2099	0.0013
$b_2$	0.8188	0.0515	15.9093	0.0000

**Raízes do polinômio AR**

solução 1	solução 2
3.0509	1.0187

**Critérios de informação**

AIC	-0.3192
BIC	-0.2772

**Testes de diagnóstico**

Teste de autocorrelação de *Ljung-Box*

	Resíduos <i>standardizados</i>		Resíduos <i>standardizados</i> ao quadrado	
	Estatística	p-value	Estatística	p-value
<i>Lag1</i>	0.2373	0.6262	5.2332	0.0222
<i>Lag2</i>	0.2478	0.8835	5.9315	0.0515
<i>Lag3</i>	2.5156	0.4725	6.8760	0.0760
<i>Lag4</i>	6.5203	0.1635	7.3538	0.1183
<i>Lag5</i>	6.5405	0.2571	8.0623	0.1528
<i>Lag6</i>	6.6761	0.3518	9.9856	0.1253
<i>Lag12</i>	12.9626	0.3718	12.7034	0.3910

Estatísticas e teste de normalidade dos resíduos *standardizados*

Curtose	4.2649
Assimetria	0.2756
<i>Jarque-Bera</i>	50.4479
(p value)	0.0000

**Especificação do modelo**

Modelo 2.b) Probabilidades de transição constantes

$$\Delta r_t = \begin{cases} a_{01} + a_{11} r_{t-1} + a_{21} r_{t-2} + (\sqrt{b_{01} + b_{11} \varepsilon_{t-1}^2 + b_{21} h_{t-1}}) z_t & \text{se } s_t = 1 \\ a_{02} + a_{12} r_{t-1} + a_{22} r_{t-2} + (\sqrt{b_{02} + b_{12} \varepsilon_{t-1}^2 + b_{22} h_{t-1}}) z_t & \text{se } s_t = 2 \end{cases}$$

$$z_t \sim i.i.d. N(0,1)$$

$$\Delta r_t | \Phi_{t-1} \sim \begin{cases} N(a_{01} + a_{11} r_{t-1} + a_{21} r_{t-2}, h_{1t}) & \text{com } p_{1t} \\ N(a_{02} + a_{12} r_{t-1} + a_{22} r_{t-2}, h_{2t}) & \text{com } 1 - p_{1t} \end{cases}$$

$$h_{it} = b_{0i} + b_{1i} \varepsilon_{t-1}^2 + b_{2i} h_{t-1}$$

**Resultados da estimação**

Mean log-likelihood 0.2052				
	estimativas	desvio-padrão	estatística t	p-value
$a_{01}$	0.1330	0.0510	2.6057	0.0092
$a_{02}$	0.0645	0.0352	1.8306	0.0672
$a_{11}$	0.5469	0.0615	8.8865	0.0000
$a_{12}$	0.0875	0.0431	2.0288	0.0425
$a_{21}$	-0.5666	0.0600	-9.4374	0.0000
$a_{22}$	-0.1019	0.0430	-2.3721	0.0177
$b_{01}$	0.0004	0.0005	0.7811	0.4347
$b_{11}$	0.1694	0.0594	2.8531	0.0043
$b_{21}$	0.8432	0.0549	15.3720	0.0000
$b_{02}$	0.0355	0.0056	6.2938	0.0000
$b_{12}$	0.2067	0.0885	2.3363	0.0195
$b_{22}$	0.1057	0.0992	1.0652	0.2868
P	0.9995	0.0005	2138.4214	0.0000
Q	0.9982	0.0008	1202.3209	0.0000

**Raízes do polinômio AR**

solução 1	solução 2
Regime 1	1.6787
Regime 2	1.0513
Regime 2	9.6552
	1.0163

**Critérios de informação**

AIC	-0.3664
BIC	-0.2683

**Testes de diagnóstico**

Teste de autocorrelação de *Ljung-Box*

	Resíduos <i>standardizados</i>		Resíduos <i>standardizados</i> ao quadrado	
	Estatística	p-value	Estatística	p-value
<i>Lag1</i>	1.2244	0.2685	1.1568	0.2821
<i>Lag2</i>	2.8108	0.2453	2.4228	0.2978
<i>Lag3</i>	3.4895	0.3221	2.4260	0.4888
<i>Lag4</i>	6.9581	0.1381	2.4407	0.6553
<i>Lag5</i>	7.4400	0.1899	2.4915	0.7778
<i>Lag6</i>	7.4900	0.2779	10.5424	0.1036
<i>Lag12</i>	11.1978	0.5120	16.5692	0.1665

Estatísticas e teste de normalidade dos resíduos *standardizados*

Curtose	3.5911
Assimetria	0.2555
<i>Jarque-Bera</i>	16.1792
(p value)	0.0003

No modelo **2.a)**, ou seja com apenas um único regime, os testes de significância individual apontam para a significância estatística de todas as estimativas dos parâmetros da média condicional (para um nível de significância de 5%). Já no caso das estimativas dos parâmetros da variância condicional parece existir evidência de insignificância estatística do termo constante,  $b_0$ . A estimativa dos parâmetros GARCH

é próxima da unidade, fixando-se em 0.9894. O valor elevado de  $b_2$  aponta para um efeito prolongado no tempo dos choques individuais.

A taxa de juro média de longo prazo implícita assume um valor semelhante ao modelo linear na média condicional analisado no ponto anterior, situando-se em 5.9%.

É possível comparar o modelo **1.a)** e **2.a)**, usando o teste  $RV^{39}$ , com o intuito de avaliar se os dois parâmetros adicionais deste último modelo, são ou não estatisticamente relevantes. Assim, neste caso teríamos como hipóteses de teste,  $H_0$ : os dois parâmetros adicionais do modelo 2.a) são nulos vs  $H_1$ : pelo menos um dos parâmetros adicionais do modelo 2.a) é diferente de zero. O valor da estatística de teste  $RV$  obtido foi de 107.67. Considerando  $\alpha=5\%$ , e tendo por base o valor crítico assintótico da distribuição  $\chi^2$ , com dois graus de liberdade, de 5.99, rejeita-se a hipótese nula, existindo, assim, evidência estatística de que o efeito conjunto dos parâmetros GARCH seja estatisticamente significativo.

À semelhança do modelo **1.a)**, os testes de diagnóstico apontam para a não rejeição da hipótese nula de ausência de autocorrelação nos resíduos estandardizados, sugerindo, assim, que a média condicional se encontra bem especificada. Já no que respeita à variância condicional, existem indícios de que este modelo, face ao modelo **1.a)**, capta melhor a heterocedasticidade presente nos dados, na medida em que as estatísticas de *Ljung-Box* dos resíduos ao quadrado estandardizados se reduziram substancialmente. No entanto, parece continuar a existir evidência de má modelação da volatilidade estocástica das taxas de juro de longo prazo (para um nível de significância de 5%) já que o teste de *Ljung-Box* sugere a presença de autocorrelação de ordem um nos resíduos estandardizados ao quadrado (para  $\alpha = 5\%$ ).

No modelo **2.b)**, que assume dois regimes e efeitos GARCH(1,1) simultaneamente, há evidência de significância estatística de praticamente todas as estimativas dos parâmetros da média condicional. De notar que apenas a estimativa de  $a_{02}$  parece ser estatisticamente insignificante, mas o valor do *p-value* está muito próximo do valor usual de referência de 5%, abaixo do qual se considera existirem indícios de significância estatística. No que respeita aos parâmetros da variância

---

<sup>39</sup> Este teste tem como estatística de teste:  $LR = 2 \times T \times [mean \log_1 - mean \log_0] \sim \chi^2_{(p_1 - p_0)}$  sob  $H_0$ , onde  $T$  representa o número de observações e  $mean \log_1$  e  $mean \log_0$  constituem o valor médio do logaritmo da verosimilhança do modelo não restrito e do modelo restrito, respectivamente.  $p_1$  é o número de parâmetros do modelo não restrito e  $p_0$  o número de parâmetros do modelo restrito.

condicional existe evidência de significância estatística da generalidade das estimativas dos parâmetros, com a exceção da constante do regime 1,  $\hat{b}_{01}$ , e da estimativa do parâmetro GARCH no regime 2,  $\hat{b}_{22}$ . Por sua vez, à semelhança do modelo não linear na média condicional, **1.b**), os parâmetros  $P$  e  $Q$  parecem ser muito persistentes, já que em ambos os regimes o valor da estimativa é superior a 0.99.

De notar ainda que em ambos os regimes as raízes estão próximas do círculo unitário, o que pode ser interpretado também como um sintoma de persistência. Os dois regimes têm raízes fora do círculo unitário, apesar de estarem próximos da não estacionaridade. Como os resultados apontam no mesmo sentido, globalmente pode-se dizer que o processo deve ser estacionário ao nível da média. Já no que respeita à questão da estacionaridade ao nível da variância, parecem existir indícios de estacionaridade no regime 2, já que a soma dos parâmetros GARCH é substancialmente inferior à unidade ( $\hat{b}_{12} + \hat{b}_{22} \approx 0.3124$ ), mas no regime 1, como a soma é próxima do valor unitário ( $\hat{b}_{11} + \hat{b}_{21} \approx 1.0127$ ), coloca-se a questão de o processo poder ser IGARCH. Neste contexto, considerou-se necessário testar a hipótese de no regime 1 a soma dos dois parâmetros não ser estatisticamente diferente de um. Assim, foi efectuado o teste de combinação linear dos parâmetros<sup>40</sup>, considerando como hipóteses de teste:  $H_0 : b_{11} + b_{21} = 1 \Leftrightarrow b_{11} + b_{21} - 1 = 0$  e  $H_1 : b_{11} + b_{21} > 1 \Leftrightarrow b_{11} + b_{21} - 1 > 0$ . Tomando como referência o nível de significância de 5%, o valor da estatística de teste (0.4224) é inferior ao valor crítico assintótico (1.645), existindo evidência a favor da não rejeição da hipótese nula. Assim, no regime 1 há indícios de o processo ser IGARCH e, portanto, não estacionário de segunda ordem, mas quase certamente estacionário em sentido estrito (ver Nelson 1990).

No regime 1, a taxa de juro média de longo prazo é de 6.7%, sendo que no regime 2 assume um valor relativamente menor, de aproximadamente 4.5%. Em termos de interpretação económica dos resultados de estimação, no regime 1, onde a taxa de juro de longo prazo é em média mais elevada, os choques individuais parecem ter um impacto pequeno, já que a estimativa do parâmetro ARCH,  $\hat{b}_{11}$ , se situa em 0.1694 e

---

<sup>40</sup> A estatística de teste  $t$  é definida como:  $t = \frac{\hat{b}_{11} + \hat{b}_{21} - 1}{\sqrt{(\text{vâr}(b_{11}) + \text{vâr}(b_{21}) + 2 \text{côv}(b_{11}, b_{21}))}}$  e, no contexto em análise,  $t$  segue uma distribuição, assintoticamente,  $N(0,1)$ , mesmo sob  $H_0$ , ou seja, sob não estacionaridade (para informação mais detalhada sobre este resultado, ver Lumsdaine (1996)).

esse efeito demora muito tempo a desaparecer (a estimativa do parâmetro GARCH,  $\hat{b}_{21}$  é de 0.8432). Por sua vez, no regime 2, em que se verifica uma taxa de juro mais reduzida, a estimativa do parâmetro ARCH,  $\hat{b}_{12}$ , de aproximadamente 0.2067, indicia que os choques individuais têm um impacto mais significativo na variância condicional, no entanto, segundo a estimativa do parâmetro GARCH, esse efeito desaparece mais rapidamente (o valor de  $\hat{b}_{22}$  é de 0.1057). Em conformidade, os regimes parecem distinguir-se entre si não só em termos de valor médio de taxa de juro, como também em termos de dinâmica de volatilidade. Adicionalmente, o regime 1 aparentemente caracteriza-se por ser um processo IGARCH, enquanto o regime 2 se define por um processo estacionário de 2º ordem. Estes resultados parecem ir ao encontro da evidência encontrada por Gray (1996) no aspecto em que este apontou para a existência de diferentes dinâmicas a caracterizar a média e a variância condicional das taxas de juro de curto prazo. Contudo, existem diferenças em termos dos aspectos que caracterizam cada regime, já que, contrariamente ao obtido por Gray, apenas um dos regimes é estacionário em covariância e os períodos de taxas de juro mais elevadas parecem ser aqueles em que os choques individuais têm um efeito imediato menor, mas mais persistente, na variância condicional.

Em relação aos testes de autocorrelação de *Ljung-Box*, estes revelaram ausência de autocorrelação nos resíduos estandardizados e nos resíduos estandardizados ao quadrado, o que parece sugerir a evidência estatística de uma adequada modelação da média e da variância condicional em simultâneo, o que não acontecia nos modelos da **classe 1**. Mais uma vez a hipótese de normalidade é rejeitada, no entanto, assistiu-se a uma redução significativa da estatística de teste *Jarque-Bera* face aos modelos da **classe 1**, o que sugere que os desvios da normalidade aparentemente são menos pronunciados, em particular, no modelo **2.b**).

Em termos globais, comparando o desempenho do modelo linear e não linear na média condicional, **2.a**) e **2.b**), parecem haver indícios de alguma superioridade deste último modelo face ao primeiro em termos de ajustamento aos dados. A incorporação de dois regimes conduz a uma diminuição significativa do critério de informação AIC e ao aumento do valor médio do logaritmo da verosimilhança, que passou de 0.1690, para 0.2052. No entanto, o valor do BIC aumentou ligeiramente. Já no teste de Davies, como se obtém  $\Pr(\sup LR > LR_{obs}) \leq 0.0000$ , significa que o verdadeiro *p-value* é inferior a

zero, e portanto, a hipótese nula é rejeitada para qualquer nível de significância. Adicionalmente, a estatística de *Ljung-Box* dos resíduos estandardizados ao quadrado sugere a ausência de autocorrelação, o que parece indicar que o modelo com alteração de regime, **2.b)**, consegue modelar adequadamente a volatilidade estocástica da taxa de juro de longo prazo. Ressalva-se, ainda, o facto de o modelo com dois regimes permitir uma caracterização diferenciada da variância condicional em cada regime. Em conformidade, considerando conjuntamente estes aspectos parece existir evidência a favor do modelo com dois regimes.

Por fim também se pode comparar ainda o modelo **2.b)** com o modelo mais simples **1.b)**, que admite igualmente a possibilidade de alteração de regime. Com efeito, pode-se efectuar o teste RV com o intuito de comparar se os 4 parâmetros adicionais do modelo **2.b)** são relevantes do ponto de vista estatístico. As hipóteses de teste poderiam ser enunciadas como:  $H_0$ : os quatro parâmetros adicionais do modelo **2.b)** são nulos vs  $H_1$ : pelo menos um dos parâmetros adicionais do modelo **2.b)** é diferente de zero. O valor da estatística de teste RV obtido foi de 19.83. Tendo por base o valor crítico assintótico de 9.49, associado à distribuição  $\chi^2$  com quatro graus de liberdade para um nível de significância de 5%, existe evidência a favor da rejeição da hipótese nula. Assim, pode-se dizer que o efeito conjunto dos parâmetros GARCH parece ser estatisticamente significativo.

De seguida irá apresentar-se os resultados obtidos para os modelos que combinam as características deste tipo de modelo, com o efeito do nível da taxa de juro na modelação da heterocedasticidade condicional.

### 3.2.3. Modelo com efeitos GARCH e o nível da taxa de juro<sup>41</sup>

Neste enquadramento serão analisados o modelo linear na média condicional e o modelo não linear com probabilidades de transição que variam no tempo, sendo que neste último caso, o modelo assume uma forma funcional que representa a formulação mais genérica apresentada por Gray na sua análise (e que designou por *generalized*

---

<sup>41</sup> À semelhança de Gray, optou-se por não incluir o termo constante na variância condicional, já que em testes realizados, verificou-se que a estimativa deste parâmetro, quando incluído, convergia para zero. De notar, que a omissão deste parâmetro da variância condicional não é problemática já que a introdução do termo relacionado com o nível da taxa de juro, equivale à incorporação de uma constante, pois esse termo actua como um limite mínimo ( $\sigma^2 \times \text{mínimo}(r_{t-1})$ ).

regime-switching model). Neste subtópico os modelos estudados serão designados como sendo da classe **3**. Em conformidade, o modelo com um regime será denominado por **3.a)** e o não linear por **3.b)**. Em qualquer um dos casos, a variância condicional será uma função de efeitos GARCH(1,1) e do nível da taxa de juro de longo prazo, assumindo, à semelhança do ponto **3.2.1.**, e pelas razões aí mencionadas,  $\gamma = 0.5$ .

Na **tabela 4** reportam-se os resultados de estimação destes dois modelos.

**Tabela 4 – Resultados de estimação do modelo 3.a) e 3.b)**

**Especificação do modelo**

Modelo 3.a)

$$\Delta r_t = a_0 + a_1 r_{t-1} + a_2 r_{t-2} + (\sqrt{b_1 \varepsilon_{t-1}^2 + b_2 h_{t-1} + \sigma_1^2 r_{t-1}^{2.05}}) z_t$$

$$z_t \sim i.i.d. N(0,1)$$

$$\Delta r_t | \Phi_{t-1} \sim N(a_0 + a_1 r_{t-1} + a_2 r_{t-2}, h_t)$$

$$h_t = b_1 \varepsilon_{t-1}^2 + b_2 h_{t-1} + \sigma_1^2 r_{t-1}$$

**Resultados da estimação**

	Mean log-likelihood		0.1682	
	estimativas	desvio-padrão	estatística t	p-value
$a_0$	0.0716	0.0338	2.1206	0.0340
$a_1$	0.3067	0.0448	6.8486	0.0000
$a_2$	-0.3187	0.0448	-7.1214	0.0000
$b_1$	0.1704	0.0533	3.1996	0.0014
$b_2$	0.8209	0.0511	16.0739	0.0000
$\sigma_1$	0.0140	0.0047	2.9660	0.0030

Raízes do polinómio AR	solução 1	solução 2
	3.0821	1.0181

**Critérios de informação**

AIC	-0.3176
BIC	-0.2755

**Testes de diagnóstico**

Teste de autocorrelação de *Ljung-Box*

	Resíduos standardizados		Resíduos standardizados ao quadrado	
	Estatística	p-value	Estatística	p-value
Lag1	0.1688	0.6811	4.7445	0.0294
Lag2	0.1702	0.9184	5.3627	0.0685
Lag3	2.6285	0.4525	6.2199	0.1014
Lag4	6.3077	0.1773	6.6936	0.1530
Lag5	6.3213	0.2762	7.4288	0.1907
Lag6	6.4395	0.3758	9.6207	0.1416
Lag12	12.9413	0.3733	12.4866	0.4074

**Estatísticas e teste normalidade Jarque-Bera**

Curtose	4.2606
Assimetria	0.2742
Jarque-Bera	50.0823
(p value)	0.0000

**Especificação do modelo**

Modelo 3.b) Probabilidades de transição que variam no tempo

$$\Delta r_t = \begin{cases} a_{01} + a_{11} r_{t-1} + a_{21} r_{t-2} + (\sqrt{b_{11} \varepsilon_{t-1}^2 + b_{21} h_{t-1} + \sigma_1^2 r_{t-1}^{2.05}}) z_t & \text{se } s_t = 1 \\ a_{02} + a_{12} r_{t-1} + a_{22} r_{t-2} + (\sqrt{b_{12} \varepsilon_{t-1}^2 + b_{22} h_{t-1} + \sigma_2^2 r_{t-1}^{2.05}}) z_t & \text{se } s_t = 2 \end{cases}$$

$$z_t \sim i.i.d. N(0,1)$$

$$\Delta r_t | \Phi_{t-1} \sim \begin{cases} N(a_{01} + a_{11} r_{t-1} + a_{21} r_{t-2}, h_t) & \text{com } p_{1t} \\ N(a_{02} + a_{12} r_{t-1} + a_{22} r_{t-2}, h_t) & \text{com } 1 - p_{1t} \end{cases}$$

$$h_t = b_{1t} \varepsilon_{t-1}^2 + b_{2t} h_{t-1} + \sigma_t^2 r_{t-1}$$

**Resultados da estimação**

	Mean log-likelihood		0.2101	
	estimativas	desvio-padrão	estatística t	p-value
$a_{01}$	0.1254	0.0478	2.6223	0.0087
$a_{02}$	0.0414	0.0405	1.0216	0.3070
$a_{11}$	0.5539	0.0572	9.6898	0.0000
$a_{12}$	0.1336	0.0584	2.2876	0.0222
$a_{21}$	-0.5722	0.0575	-9.9515	0.0000
$a_{22}$	-0.1418	0.0582	-2.4361	0.0148
$b_{11}$	0.0742	0.1248	0.5944	0.5523
$b_{21}$	0.6316	0.1441	4.3831	0.0000
$\sigma_1$	0.0225	0.0051	4.3808	0.0000
$b_{12}$	0.1762	0.0913	1.9300	0.0536
$b_{22}$	0.2574	0.5252	0.4902	0.6240
$\sigma_2$	0.0780	0.0316	2.4650	0.0137
$c_1$	3.1023	0.8183	3.7913	0.0001
$d_1$	-0.0974	0.1069	-0.9113	0.3622
$c_2$	-3.3945	9.0698	-0.3743	0.7082
$d_2$	2.6265	2.9891	0.8787	0.3796

Raízes do polinómio AR	solução 1	solução 2
Regime 1	1.6677	1.0479
Regime 2	6.9824	1.0097

**Critérios de informação**

AIC	-0.3698
BIC	-0.2577

**Testes de diagnóstico**

Teste de autocorrelação de *Ljung-Box*

	Resíduos standardizados		Resíduos standardizados ao quadrado	
	Estatística	p-value	Estatística	p-value
Lag1	0.1437	0.7046	0.0074	0.9314
Lag2	1.0591	0.5889	0.2718	0.8729
Lag3	1.7302	0.6302	0.6967	0.8740
Lag4	5.2567	0.2620	1.8440	0.7644
Lag5	5.2764	0.3831	2.1212	0.8321
Lag6	5.2810	0.5083	5.4583	0.4865
Lag12	8.7122	0.7273	9.4304	0.6658

**Estatísticas e teste de normalidade dos resíduos standardizados**

Curtose	3.8450
Assimetria	0.3863
Jarque-Bera	34.7442
(p value)	0.0000

No modelo em que se assume apenas um único regime, **3.a**), existe evidência de significância estatística de todas as estimativas dos parâmetros (para um nível de significância de 5%). À semelhança do modelo **2.a**) (modelo linear na média condicional analisado no ponto anterior), o parâmetro GARCH associado à persistência,  $b_2$ , assume um valor elevado, o que confirma o resultado já evidenciado nesse modelo que apontava para o efeito relativamente prolongado dos choques individuais. Em conformidade, a soma da estimativa dos parâmetros GARCH é próxima da unidade (0.9914). A estimativa do parâmetro associada ao nível das taxas de juro,  $\sigma_1$ , revelou-se estatisticamente significativa.

A taxa de juro média de longo prazo implícita também assume um valor análogo ao modelo **2.a**) e **1.a**), de 6.0%.

Em termos de testes de diagnóstico, os resultados obtidos para o modelo **3.a**) sugerem que este sofre do mesmo tipo de problemas evidenciados pelo modelo de um regime analisado nos dois pontos anteriores, **2.a**) e **1.a**) (ou seja, indícios de inadequada especificação da variância condicional). De notar, ainda, que os modelos **3.a**) e **2.a**) têm o mesmo número de parâmetros e assistiu-se a uma ligeira redução do valor do logaritmo de verosimilhança (0.1682 vs 0.1690, respectivamente).

Já no modelo **3.b**), em que se admite a presença de dois regimes com efeitos GARCH(1,1) e o nível da taxa de juro na modelação da variância condicional num contexto em que as probabilidades de transição variam no tempo, as estimativas da média condicional, são genericamente semelhantes, em termos de dimensão e significância estatística às do modelo anterior, **2.b**) (modelo que só assumia a presença de efeitos GARCH e probabilidades de transição constantes). Em conformidade, a taxa de juro média de longo prazo implícita, no regime 1, é de 6.8% e no regime 2 de 5.0%. Já em termos de variância condicional, existe evidência de insignificância estatística da estimativa do parâmetro ARCH do regime 1,  $\hat{b}_{11}$ , e do parâmetro GARCH do regime 2,  $\hat{b}_{22}$ . Em termos de interpretação económica e dimensão das estimativas, à semelhança do modelo **2.b**), o regime 1 é aquele que aparentemente regista uma taxa de juro média mais elevada, e os choques individuais parecem ter um impacto relativamente mais pequeno (0.0742) e com maior persistência (0.6316) comparativamente ao regime 2 (de 0.1762 e 0.2574, respectivamente). Assim, tal como aparentemente se verificava no **2.b**) e **1.b**), os regimes parecem distinguir-se entre si pelo valor médio de taxa de juro, bem

como em termos de dinâmica de volatilidade. Por sua vez, como as raízes do polinómio estão fora do círculo unitário nos dois regimes, ainda que assumam valores próximos da não estacionaridade, globalmente pode-se dizer que o processo deve ser estacionário ao nível da média. Existe, contudo, uma diferença assinalável entre o modelo **3.b)** e **2.b)**, que se prende com o facto de o primeiro, contrariamente ao segundo, se caracterizar pela presença de indícios que apontam para a estacionaridade em covariância dos dois regimes ao nível da variância condicional, já que a soma dos parâmetros GARCH é substancialmente inferior à unidade quer no regime 1 quer no regime 2. Assim, à semelhança de Gray (1996), os dois regimes parecem ser estacionários em covariância e a persistência dos choques individuais é menor do que o que sugere o modelo com apenas um regime, **3.a)**<sup>42</sup>. Este resultado parece confirmar a evidência encontrada por Gray que apontava para a incapacidade do modelo com apenas um regime para capturar a persistência da volatilidade que advém do facto de existirem regimes persistentes e, nesse contexto, no modelo com um regime toda a persistência da volatilidade é “capturada” pelo coeficiente do parâmetro GARCH (que mede a persistência dos choques individuais). Já no modelo com dois regimes, uma parte da persistência da volatilidade é capturada pelos parâmetros associados aos regimes (dos quais  $P_t$  e  $Q_t$  são uma função) e, portanto, as estimativas associadas à persistência dentro de cada regime,  $\hat{b}_{22}$  e  $\hat{b}_{21}$ , tendem assumir valores mais reduzidos. No entanto, é de ressaltar que tal como no modelo **2.b)** e contrariamente ao obtido por Gray para as taxas de juro de curto prazo, existe evidência de que os períodos de taxas de juro de longo prazo mais elevadas se caracterizam por um efeito imediato menor dos choques individuais na variância condicional, e mais persistente.

A probabilidade de permanecer no regime 1,  $P_t$ , é determinada pelos coeficientes  $c_1$  e  $d_1$ , enquanto que a probabilidade de continuar no segundo regime,  $Q_t$ , é uma função de  $c_2$  e  $d_2$ . É de notar que o desvio padrão das estimativas destes parâmetros é relativamente elevado<sup>43</sup>, nomeadamente de  $c_2$  e  $d_2$ , o que parece sugerir alguma imprecisão na sua estimação. Adicionalmente, em termos de interpretação de resultados, aparentemente a probabilidade de se estar num determinado regime e permanecer nele é

---

<sup>42</sup> No modelo **3.a)**  $\hat{b}_1 + \hat{b}_2 \approx 0.9914$  enquanto que no modelo **3.b)**, no regime 1  $\hat{b}_{11} + \hat{b}_{21} \approx 0.7058$  e no regime 2  $\hat{b}_{12} + \hat{b}_{22} \approx 0.4337$ .

<sup>43</sup> Este facto também se encontrava presente nos resultados encontrados por Gray (1996).



elevada e relativamente insensível ao nível da taxa de juro. Apesar de  $d_1$  ser negativo, para que a subida da taxa de juro conduzisse a uma redução expressiva da probabilidade de estar no regime 1, seria necessário que a taxa de juro atingisse valores muito superiores aos observados na amostra. Tendo em conta o valor da estimativa do termo  $c_1$  e que o nível da taxa de juro na amostra oscila entre o mínimo de 2.9% e o máximo de 11.4%, a probabilidade de continuar nesse regime é muito elevada em qualquer um dos casos (sempre superior a 0.97%), revelando, por isso, pouca sensibilidade ao nível da taxa de juro. Por sua vez, no regime 2, onde a taxa de juro assume um valor médio mais reduzido e a estimativa de  $d_2$  é positiva, mesmo quando a taxa de juro assume o valor mais baixo da série (2.9%) a probabilidade de continuar nesse regime é muito elevada (próxima da unidade) e, portanto, para níveis superiores da taxa de juro o efeito positivo da subida da taxa de juro sobre a probabilidade de continuar nesse regime é muito diminuto e, por isso, sem expressão. Para que a subida da taxa de juro se traduzisse num aumento significativo da probabilidade de continuar no regime 2, seria necessário que a taxa de juro se situasse em níveis muito inferiores aos observados no período amostral. Para além do desvio padrão das estimativas dos parâmetros  $c_2$  e  $d_2$  ser relativamente elevado, ressalva-se ainda o facto de em várias experiências realizadas se ter constatado que diferentes valores iniciais para os parâmetros conduziam a um valor semelhante da função logaritmo de verosimilhança mas com estimativas ligeiramente diferentes para os parâmetros  $c_2$  e  $d_2$ . O modelo seleccionado foi o apresentado na **tabela 4**, por ser o que apresentou o valor mais baixo em termos de critérios de informação (AIC/BIC).

De notar ainda que no modelo **3.b**), os testes de autocorrelação de *Ljung-Box* apontaram para a não rejeição da hipótese nula nos resíduos standardizados e nos resíduos standardizados ao quadrado, o que indicia uma adequada modelação da média e da variância condicional para os níveis usuais de significância. A hipótese de normalidade é rejeitada, tendo-se verificado um incremento da estatística de teste *Jarque-Bera*, face ao modelo que incorporava apenas efeitos GARCH, **2.b**). Tal facto indica que os desvios da normalidade aparentemente são mais pronunciados no modelo com dois regimes da classe **3**.

Em termos de ajustamento aos dados, comparando o desempenho do modelo com um regime e com dois regimes, **3.a**) e **3.b**), existem indícios que favorecem a escolha por este último. Se por um lado, o modelo **3.a**) apresenta um valor inferior do

critério de informação BIC, a introdução de dois regimes, no modelo **3.b**), resultou numa diminuição significativa do critério AIC. De acordo com o teste de Davies  $\Pr(\sup LR > LR_{obs}) \leq 0.0000$ , rejeitando-se a hipótese nula de um regime para qualquer nível de significância. A estatística de *Ljung-Box* dos resíduos estandardizados ao quadrado, tal como acontecia no modelo **1.a**) e **2.a**), indicia que o modelo com apenas um regime não consegue modelar adequadamente a volatilidade estocástica da taxa de juro de longo prazo. Por fim, tal como já indicia o modelo só com efeitos GARCH, **2.b**), a taxa de juro de longo prazo parece ter um comportamento da média e de variância diferenciados em cada regime, o que só pode ser captado pela incorporação de dois regimes.

Reconhecendo a importância dos modelos com dois regimes, faz sentido efectuar o teste do RV com o objectivo de comparar o modelo **3.b**) (dois regimes com efeitos GARCH e nível da taxa de juro com probabilidades de transição que variam no tempo) com o modelo anterior **2.b**) (dois regimes com apenas efeitos GARCH e com probabilidades de transição constantes), relativamente mais simples. Com efeito, definindo as hipóteses de teste como  $H_0: d_1 = d_2 = 0$  e  $H_1: \exists d_i \neq 0, i = 1, 2$ , o valor da estatística de teste LR obtido foi de 6.18. Como RV segue uma distribuição  $\chi^2_2$  sob  $H_0$ , para  $\alpha = 5\%$  (o valor crítico assintótico é de 5.99) existe evidência a favor da rejeição da hipótese nula, contudo, o valor está muito próximo da zona de não rejeição. Considerando  $\alpha = 1\%$  (o valor crítico assintótico é de 9.21) há indícios que apontam para a não rejeição da hipótese nula. Ou seja, neste caso os resultados do teste de RV não revelam indícios claros sobre a rejeição ou não da hipótese nula.

### 3.3. Selecção dos modelos e ajustamento *in sample*

Os resultados de estimação obtidos nas diversas especificações consideradas, e apresentadas no ponto anterior, sugerem a presença de heterocedasticidade condicional, com uma dinâmica compatível com a formulação GARCH. Os modelos da **classe 1** revelaram-se incapazes de captar a dinâmica da variância condicional, existindo indícios da presença de autocorrelação nos resíduos estandardizados ao quadrado. Já nos modelos da **classe 2** verificou-se uma redução significativa da estatística de teste de

*Ljung-Box* o que sugere uma maior capacidade para captar a heterocedasticidade presente nos dados, sendo que no caso do modelo que incorporava a hipótese de dois regimes, **2.b)**, os testes de diagnóstico apontam para uma adequada especificação da variância condicional. Adicionalmente, os critérios de informação diminuíram<sup>44</sup> e o teste do RV revelou indícios de significância estatística do efeito conjunto dos parâmetros GARCH. Por fim, foram analisados os modelos da classe **3**, que incorporam uma complexidade maior relativamente aos da classe **2**, ao introduzir-se o efeito nível na modelação da variância condicional e a possibilidade das probabilidades de transição variarem no tempo. No caso do modelo com um regime valores mais reduzidos dos critérios AIC/BIC e mais elevados do valor da função logaritmo de verosimilhança favorecem a formulação mais simples, que assume apenas efeitos GARCH na modelação da variância condicional. Já no caso do modelo não linear os resultados revelaram-se um pouco contraditórios, já que se assistiu a um decréscimo ligeiro do critério AIC e a um aumento do BIC no modelo **3.b)** face ao modelo **2.b)**. Por sua vez, o teste do RV quando se assume  $\alpha = 5\%$  parece existir evidência a favor do modelo mais complexo, **3.b)** mas quando se considera  $\alpha = 1\%$  surgem indícios que favorecem a escolha pelo modelo **2.b)**. Adicionalmente, a este respeito é de notar que apesar do modelo **3.b)** ter dois parâmetros adicionais, que procuram captar o efeito do nível da taxa de juro sobre a probabilidade de permanecer num determinado regime,  $d_1$  e  $d_2$ , o valor das estimativas obtido parece sugerir alguma insensibilidade dessas probabilidades, pelo menos na amostra considerada, ao nível da taxa de juro. Já a estatística de *Jarque-Bera* parece sugerir a presença de desvios menos pronunciados face à normalidade no modelo **2.b)**. Tendo em conta estes aspectos, a par do facto do modelo **2.b)** ser mais parcimonioso e, aparentemente, mais robusto em termos de estimação, optou-se por escolher os modelos da classe **2** para prosseguir análise em termos de avaliação da performance do modelo linear e não linear na média condicional no que respeita ao ajustamento aos dados *in sample* e também em termos de previsão *out of sample*.

Após a escolha da classe de modelos a ter em consideração (classe **2**), decidiu-se eliminar os parâmetros estatisticamente insignificantes, e reestimar novamente os modelos. No caso do modelo **2.a)**, o único parâmetro estatisticamente insignificante é a estimativa da constante da formulação da variância, e, por isso, optou-se por mantê-lo já

---

<sup>44</sup> Com a excepção do critério BIC no caso do modelo **2.b)** que registou um ligeiro acréscimo face ao modelo **1.b)**.

que a ausência deste parâmetro significaria que teoricamente a variância seria nula. Assim, o modelo considerado doravante mantém-se tal e qual como apresentado na **tabela 3**. Já no caso do modelo **2.b)**, e considerando um nível de significância de 5%, existem três parâmetros cujas estatísticas se revelaram estatisticamente insignificantes,  $a_{02}$ ,  $b_{01}$  e  $b_{22}$ . Como  $b_{01}$  corresponde à constante da variância, e pelo mesmo motivo acima referido para o modelo **2.a)**, optou-se pela sua não exclusão do modelo. Já no que se refere a  $a_{02}$ , como é a constante da média e a estimativa encontra-se muito próxima da significância estatística, também se optou por manter este parâmetro no modelo. Com efeito, reestimou-se o modelo **2.b)**, excluindo apenas o parâmetro GARCH do segundo regime,  $b_{22}$ , apresentando-se os resultados da reestimação na **tabela 5**.

**Tabela 5 – Resultados de estimação do modelo 2.b)\***

**Especificação do modelo**

Modelo 2.b)\* Probabilidades de transição constantes

$$\Delta r_t = \begin{cases} a_{01} + a_{11}r_{t-1} + a_{21}r_{t-2} + (\sqrt{b_{01} + b_{11}\varepsilon_{t-1}^2 + b_{21}h_{t-1}})z_t & \text{se } s_t = 1 \\ a_{02} + a_{12}r_{t-1} + a_{22}r_{t-2} + (\sqrt{b_{02} + b_{12}\varepsilon_{t-1}^2})z_t & \text{se } s_t = 2 \end{cases}$$

$z_t \sim i.i.d. N(0,1)$

$$\Delta r_t | \Phi_{t-1} \sim \begin{cases} N(a_{01} + a_{11}r_{t-1} + a_{21}r_{t-2}, b_{01} + b_{11}\varepsilon_{t-1}^2 + b_{21}h_{t-1}) & \text{com } p_t \\ N(a_{02} + a_{12}r_{t-1} + a_{22}r_{t-2}, b_{02} + b_{12}\varepsilon_{t-1}^2) & \text{com } 1-p_t \end{cases}$$

**Raízes do polinómio AR**

	solução 1	solução 2
Regime 1	1.6753	1.0520
Regime 2	9.7463	1.0169

**Critérios de informação**

AIC	-0.3680
BIC	-0.2769

**Testes de diagnóstico**

Teste de autocorrelação de Ljung-Box

	Resíduos <i>standardizados</i>		Resíduos <i>standardizados</i> ao quadrado	
	Estatística	p-value	Estatística	p-value
Lag1	1.5150	0.2184	1.0936	0.2957
Lag2	3.2300	0.1989	2.1418	0.3427
Lag3	3.8752	0.2753	2.1438	0.5431
Lag4	7.4593	0.1135	2.1793	0.7028
Lag5	7.9142	0.1610	2.2242	0.8173
Lag6	7.9652	0.2407	9.9019	0.1288
Lag12	11.4496	0.4908	16.2087	0.1819

Estatísticas e teste de normalidade dos resíduos *standardizados*

Curtose	3.5740
Assimetria	0.2464
Jarque-Bera	15.1694
(p value)	0.0005

**Resultados da estimação**

Mean log-likelihood		0.2044		
	estimativas	desvio-padrão	estatística t	p-value
$a_{01}$	0.1341	0.0522	2.5678	0.0102
$a_{02}$	0.0663	0.0413	1.6035	0.1088
$a_{11}$	0.5474	0.0765	7.1559	0.0000
$a_{12}$	0.0859	0.0403	2.1315	0.0330
$a_{21}$	-0.5674	0.0743	-7.6339	0.0000
$a_{22}$	-0.1009	0.0403	-2.5016	0.0124
$b_{01}$	0.0004	0.0006	0.7605	0.4469
$b_{11}$	0.1735	0.0626	2.7724	0.0056
$b_{21}$	0.8354	0.0613	13.6260	0.0000
$b_{02}$	0.0410	0.0037	11.0572	0.0000
$b_{12}$	0.2171	0.0724	2.9991	0.0027
P	0.9994	0.0007	1466.1061	0.0000
Q	0.9980	0.0010	1043.7999	0.0000

Face ao modelo sem imposição de restrições, **2.b)**, o critério AIC e BIC registaram um ligeiro decréscimo. A dimensão das estimativas dos parâmetros são muito semelhantes, pelo que em termos de interpretação, globalmente, mantêm-se a análise que já foi efectuada no ponto **3.2.2.**. Em conformidade, no **regime 1**, a taxa de juro média de longo prazo é mais elevada, 6.7% e no que concerne à volatilidade, este regime caracteriza-se pelo facto do efeito dos choques individuais ser relativamente pequeno, 0.1735, mas persistente (0.8354), demorando muito tempo a desaparecer no

tempo. Já no **regime 2**, a taxa de juro média é menor (4.4%) e o efeito dos choques individuais é relativamente maior (0.2171), mas pouco persistente. Em termos de estacionaridade também se mantêm as mesmas conclusões. Assim, ao nível da média como as raízes do polinómio AR estão fora do círculo unitário em ambos os regimes, pode-se dizer que, globalmente, o processo deverá ser estacionário. De notar que, como as raízes estão próximas da unidade, este é um sintoma de elevada persistência, o que significa que períodos de elevadas (reduzidas) taxas de juro tendem a ser seguidos por períodos de elevadas (reduzidas) taxas de juro. Em relação à variância como a soma da estimativa dos parâmetros GARCH no regime 1 continua a ser superior à unidade ( $\hat{b}_{11} + \hat{b}_{21} \approx 1.0089$ ), efectuou-se novamente o teste de combinação linear dos parâmetros, nos mesmos moldes enunciados anteriormente. Neste caso o valor da estatística de teste obtido foi de 0.2316, que comparando com o valor crítico assintótico ( $\alpha = 5\%$ ), de 1.645, conduz à não rejeição da hipótese nula. Assim, tal como já acontecia no modelo **2.b**), ainda que possa não existir estacionaridade de segunda ordem, quase certamente existe estacionaridade em sentido estrito.

Efectuando a comparação entre o modelo linear na média condicional, **2.a**), e não linear **2.b**)\* (ou seja, o modelo **2.b**) com a restrição de nulidade imposta sobre  $b_{22}$ ), considerando o critério de informação AIC, os testes de diagnóstico, a par dos resultados do teste de Davies<sup>45</sup>, existem indícios favorecem o modelo com dois regimes face ao modelo com um regime<sup>46</sup>. Com efeito, globalmente parece existir evidência de que o modelo não linear produz um melhor ajustamento *in sample* face ao linear.

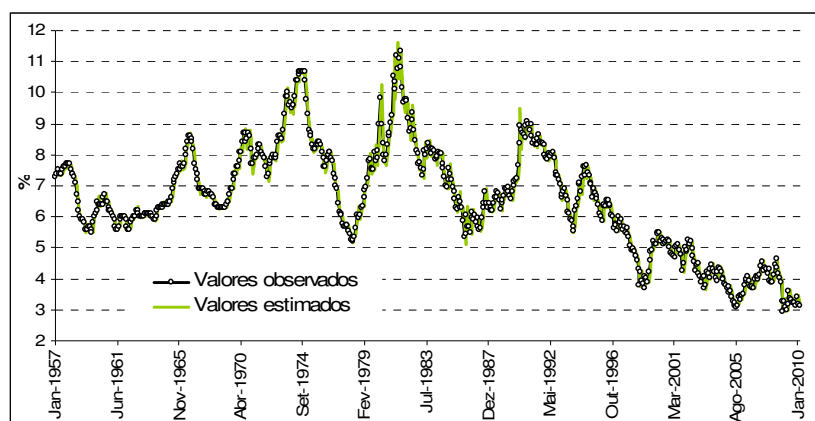
No gráfico **3** representa-se o valor observado da série conjuntamente com o valor estimado ( $E[\hat{r}_t | \Phi_{t-1}]$ ), no gráfico **4** o comportamento da média condicional e no gráfico **5** da variância condicional, do modelo com dois regimes, **2.b**)\*.

---

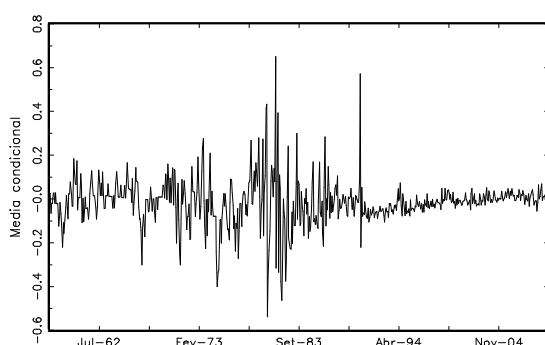
<sup>45</sup> Neste caso, o resultado do teste obtido é análogo ao anterior, ou seja,  $\Pr(\sup LR > LR_{\alpha}) \leq 0.0000$  e, portanto, a hipótese nula de 1 regime é rejeitada para qualquer nível de significância.

<sup>46</sup> Apenas o critério BIC parece apontar noutro sentido, já que é ligeiramente superior.

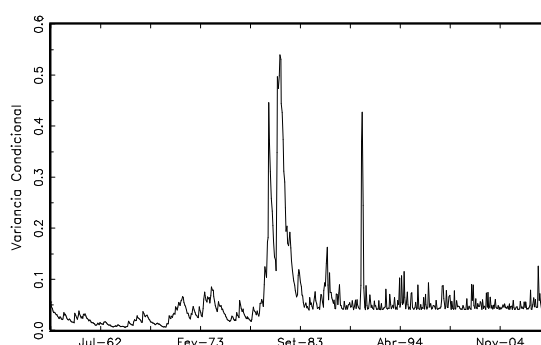
**Gráfico 3 – Yield do Bund a 10 anos – valores estimados vs observados**



**Gráfico 4 – Média condicional**

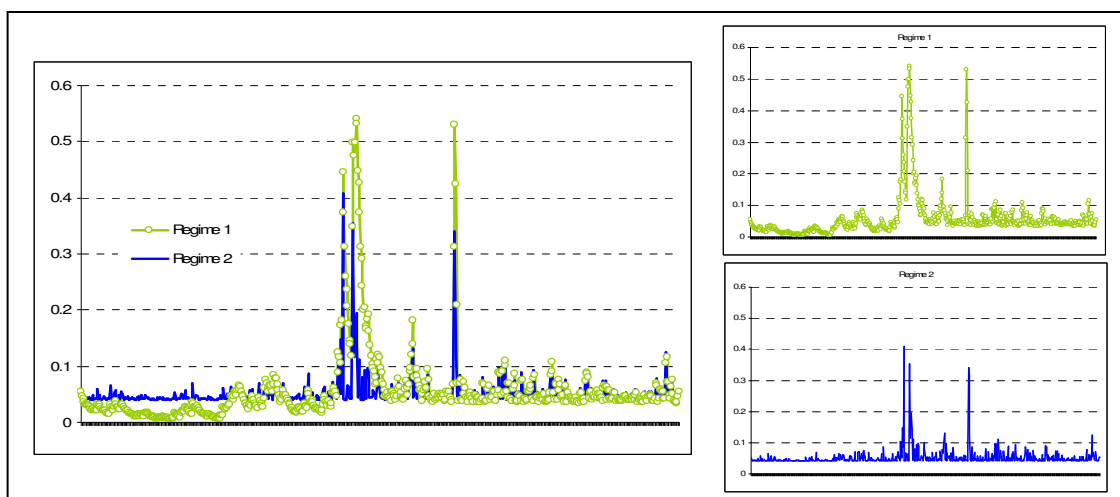


**Gráfico 5 – Variância condicional**



A análise do gráfico 3, permite concluir que o modelo não linear na média condicional, com efeitos GARCH, consegue replicar de forma razoável o comportamento histórico da série em estudo. O gráfico 4 sugere que os últimos anos têm sido marcados por oscilações menos pronunciadas comparativamente ao início da amostra e, sobretudo, face ao final da década de 70 e início dos anos 80. Por sua vez, em termos de variância (gráfico 5) parece que ocorreu uma alteração estrutural da dinâmica da volatilidade. Aparentemente, há um período inicial que se caracteriza por uma elevada persistência, ou seja, valores elevados/baixos tendem a surgir em sequência. Posteriormente, segue-se um segundo período, marcado por outro tipo de dinâmica, em que a volatilidade se caracteriza, por pouca persistência, já que parece reverter rapidamente para um valor de tendência central. Como no regime 1, o processo se caracteriza por ser um processo IGARCH, não tem variância (incondicional) porque é infinita, não sendo, por isso, estacionário em covariância. De notar que apenas existe quase certamente estacionaridade em sentido estrito e esta não exige a existência de variância finita. Neste contexto, considerou-se oportuno calcular a variância (condicional) estimada em cada regime, cujo comportamento se apresenta no gráfico 6.

Gráfico 6 – Variância empírica



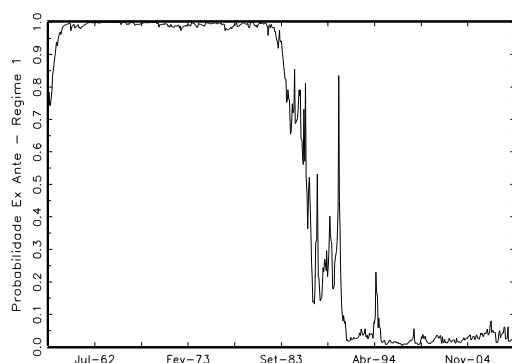
Do lado esquerdo, tem-se uma representação da variância empírica do regime 1 e 2, e do lado direito apresenta-se a mesma informação mas de forma individualizada, ou seja, no canto superior direito tem-se a variância empírica do regime 1 e no inferior direito a do regime 2. Em termos empíricos, o valor médio da variância condicional no regime 1 é de 0.057 e no regime 2 é de 0.053, pelo que aparentemente a volatilidade tende a assumir um valor médio análogo nos dois regimes e, portanto, o que os distingue parece ser a sua dinâmica, que aparentemente terá sofrido uma alteração estrutural.

Na estimação do modelo com dois regimes, há duas probabilidades condicionais importantes. A **probabilidade *ex-ante***, que pode ser entendida como  $\Pr[s_t = i | \Phi_{t-1}]$  e é pertinente para efeitos de previsão, e a **probabilidade *smoothed***, que tem em conta não apenas a informação disponível até ao momento  $t-1$ , mas sim todo o conjunto informativo  $T$ , sendo definida como  $\Pr[s_t = i | \Phi_T]$ . Esta segunda abordagem é relevante na determinação dos momentos em que ocorrem alterações de regime. Gray (1995) desenvolveu um filtro eficiente que recorre à natureza recursiva das probabilidades de estar num regime e estabelece uma relação directa entre a probabilidade *ex-ante* e a probabilidade *smoothed*<sup>47</sup>, tendo sido essa abordagem seguida neste trabalho. Contrariamente ao algoritmo desenvolvido por Kim (1994), que inicia o processo com  $\Pr[s_T = 1 | \Phi_T]$  e vai iterando progressivamente para “trás” até se obter  $\Pr[s_t = 1 | \Phi_{t-1}]$ , o algoritmo de Gray é *forward looking*, relacionando directamente a probabilidade *ex-ante* e a *smoothed probability*. Com efeito, após a obtenção da probabilidade *ex-ante*,

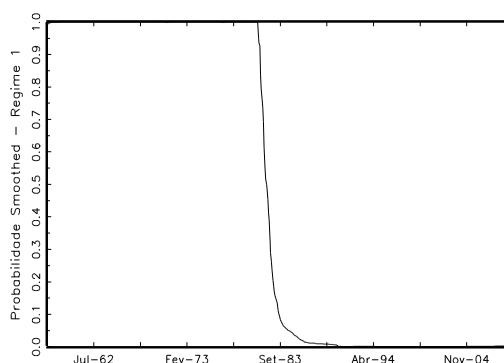
<sup>47</sup> Segundo Gray (1996), esta representação eficiente recursiva de construir a *smoothed probability* produz os mesmos resultados que o procedimento desenvolvido por Hamilton (1989, 1990).

$\Pr[s_t = 1 | \Phi_{t-1}]$ , de forma recursiva obtém-se  $\Pr[s_t = 1 | \Phi_t]$ , prosseguindo-se sucessivamente até se obter a *smoothed probability*, ou seja,  $\Pr[s_t = 1 | \Phi_T]$ . O **gráfico 7** representa a probabilidade *ex-ante* de estar no regime 1, ou seja,  $\Pr[s_t = 1 | \Phi_{t-1}]$  e o **gráfico 8** a probabilidade *smoothed*.

**Gráfico 7 – Probabilidade *ex-ante***



**Gráfico 8 – Probabilidade *smoothed***



Aparentemente o processo esteve no regime 1, de taxas de juro mais elevadas e elevada persistência dos choques individuais, com uma probabilidade significativa até ao início de 1981. A partir do final de 1987, mais concretamente após a ocorrência do *crash* no mercado accionista (Outubro de 1987), parece existir uma elevada probabilidade de o processo se encontrar no regime 2, que se caracteriza por taxas de juro de longo prazo mais baixas e com choques individuais significativamente menos persistentes, mas com efeito imediato maior face ao regime 1. De notar que este segundo regime engloba todo o período após a reunificação alemã. Já durante grande parte da década de 80, mais concretamente, entre o início de 1981 e até ao final de 1987 parece que o processo não esteve totalmente em nenhum dos dois regimes.

No próximo subcapítulo, compara-se os dois modelos aqui analisados, procurando agora avaliar a sua capacidade de previsão (*out of sample*), para o horizonte de um, dois e três meses, considerando os últimos 3 anos de observações.



### 3.4. Previsão *out of sample*

Um dos objectivos do trabalho é comparar a qualidade de previsão *out of sample* do nível da taxa de juro dos modelos do tipo linear e não linear na média condicional e, por isso, decidiu-se comparar os resultados previsão, numa frequência mensal, até ao horizonte temporal de um trimestre, ou seja, previsão a um, dois e três passos dos dois modelos seleccionados no ponto anterior, **2.a)** e **2.b)\***. Neste contexto, considerando que se tem observações disponíveis desde o momento 1 até  $T$ , dividiu-se o período de observações em dois sub-períodos. O primeiro é o período de estimação, também denominado por *in sample*, que engloba uma série de observações, por exemplo de 1 a  $t_1$ , e que será usado apenas para estimar o modelo. O segundo período é designado por *out of sample*, que inclui as observações remanescentes, ou seja, de  $t_{1+1}$  a  $T$  e serve apenas para avaliar a qualidade de previsão do modelo que foi estimado *in sample*, através da comparação das previsões com os valores observados da série, e os quais não foram utilizados na estimação. Neste trabalho optou-se por usar abordagem de “janela móvel” avaliando-se a capacidade previsão dos modelos a um, dois e três passos ao longo dos últimos 3 anos de observações, ou seja, 36 meses. Em conformidade, estimaram-se os modelos tomando em consideração o conjunto de observação disponíveis desde o momento 1 ao 600, e efectuou-se previsão para o valor da taxa de juro no momento 601, 602 e 603, que corresponde à previsão a um, dois e três passos, respectivamente. De seguida estimaram-se novamente os modelos tomando em conta as observação disponíveis desde o momento 2 ao 601, e prevê-se qual o valor da taxa de juro no momento 602, 603 e 604. Repete-se este procedimento em 36 iterações, até que por fim se estimam os modelos entre o período 36 e o 635, e efectua-se previsão para o momento 636, 637 e 638. Em qualquer uma das iterações o período de estimação contém sempre 600 observações disponíveis e, portanto, o esforço de estimação é sempre igual. No final, tem-se uma amostra com os valores observados *out of sample* e os previstos a um, dois e três passos, com o modelo **2.a)** e **2.b)\***, que podem ser, assim, comparados usando os critérios do erro médio<sup>48</sup>, REQM<sup>49</sup> e EAM<sup>50</sup>.

---

<sup>48</sup> O erro médio foi calculado simplesmente como a média do erro de cada período, ou seja,  $\frac{1}{T} \sum_{t=1}^T (r_t - \hat{r}_t)$ .

<sup>49</sup> A medida REQM (raiz do erro quadrático médio) é obtida da seguinte forma:  $\sqrt{\frac{1}{T} \sum_{t=1}^T (r_t - \hat{r}_t)^2}$ .

No caso do modelo linear, pode-se demonstrar que a previsão a  $h$  passos no contexto do modelo do tipo AR(2), pode ser facilmente obtida de forma recursiva, admitindo que se conhecem os parâmetros do modelo. Basicamente a função  $E(\Delta r_{t+h} | \Phi_t)$  é usada como previsor para  $\Delta r_{t+h}$ . Formalmente e considerando o caso genérico de previsão a  $h$  passos efectuada no momento  $t$ , tem-se que:

$$\Delta r_{t+h} = a_0 + a_1 r_{t+h-1} + a_2 r_{t+h-2} . \quad (40)$$

No modelo com dois regimes, em que o processo é representado pela formulação **MSW** a previsão é mais complexa mas, ainda assim, é relativamente simples obter uma expressão analítica para a previsão a  $h$  passos. Basicamente a previsão do valor futuro da variável,  $\Delta r_{t+h}$ , pode ser decomposta em duas previsões: **1)** a previsão de  $\Delta r_{t+h}$  condicional no regime que se irá realizar em  $t+h$ , ou seja,  $s_{t+h}$ , e **2)** a previsão da probabilidade com que cada regime deverá ocorrer no momento  $t+h$ . Neste caso e usando a função  $E(\Delta r_{t+h} | \Phi_t)$  como previsor para  $\Delta r_{t+h}$ , pode-se provar que, em termos gerais, a previsão a  $h$  passos no momento  $t$  é dada por:

$$E(\Delta r_{t+h} | \Phi_t) = \sum_{i=1}^2 E(\Delta r_{t+h} | \Phi_t, s_{t+h} = i) \Pr(s_{t+h} = i | \Phi_t) \quad (41)$$

onde  $E(\Delta r_{t+h} | \Phi_t, s_{t+h} = i)$  representa a média condicional do regime  $i$  e  $\Pr(s_{t+h} = i | \Phi_t)$  a probabilidade de se estar no regime  $i$  em  $t+h$  condicional na informação disponível até ao momento  $t$ . Assumindo que  $s$  segue uma cadeia *Markov*, a probabilidade de se estar no regime  $i$  em  $t+h$ , condicional na informação do momento  $t$  pode ser formulada como:

$$\Pr(s_{t+h} = i | \Phi_t) = \sum_{j=1}^2 \Pr(s_{t+h} = i | s_t = j) \Pr(s_t = j | \Phi_t) \quad (42)$$

---

<sup>50</sup> O EAM (erro absoluto médio) é calculado como  $\frac{1}{T} \sum_{t=1}^T |(r_t - \hat{r}_t)|$ .

onde  $\Pr(s_{t+h} = i | s_t = j)$  é o elemento  $(j, i)$ , da matriz  $M^h$  e representa a probabilidade de  $s_{t+h}$  estar no regime  $i$ , condicional que no momento  $t$  estava no regime  $j$ . De notar que  $M^h$  representa a matriz de probabilidades de transição  $h$  períodos à frente e é obtida multiplicando a matriz  $M$ , já referida no ponto 3.2. por ela própria,  $h$  vezes. Por sua vez,  $\Pr(s_t = j | \Phi_t)$ , denota a probabilidade de se estar no regime  $j$  no momento  $t$ , tendo por base a informação disponível até ao momento  $t$ , e tendo em conta a formulação (35), pode ser representada como:

$$\Pr(s_t = j | \Phi_t) = \frac{f(\Delta r_t | s_t = j, \Phi_{t-1}) \Pr(s_t = j | \Phi_{t-1})}{\sum_{i=1}^2 f(\Delta r_t | s_t = i, \Phi_{t-1}) \Pr(s_t = i | \Phi_{t-1})} . \quad (43)$$

Os resultados obtidos para os dois modelos, **2.a)** e **2.b)\***, em termos de critérios erro médio, REQM e EAM, na previsão a um, dois e três passos usando a metodologia de “janela rolante”, são apresentados na **tabela 6**.

**Tabela 6 – Previsão *out of sample* – critérios de avaliação**

	1 passo		2 passos		3 passos	
	Modelo 2.a)	Modelo 2.b)*	Modelo 2.a)	Modelo 2.b)*	Modelo 2.a)	Modelo 2.b)*
<b>Erro médio</b>	-0.0411	-0.0300	-0.1038	-0.0737	-0.1687	-0.1169
<b>REQM</b>	0.2224	0.2215	0.3499	0.3347	0.4258	0.4043
<b>EAM</b>	0.1692	0.1774	0.2727	0.2533	0.3347	0.3149

Como era de esperar à medida que aumenta o horizonte de previsão, maior é o erro cometido em ambos os modelos. Qualquer um dos critérios indiciam uma melhor capacidade de previsão *out of sample* do modelo não linear na média condicional, que tende a tornar-se mais evidente à medida que se estende o horizonte de previsão.

A avaliação da capacidade de ajustamento aos dados (*in sample*) e de previsão (*out of sample*) dos modelos lineares e não lineares na média condicional tem sido muito discutida na literatura. Se por um lado, frequentemente, os modelos com alteração de regime são referidos como tendo a vantagem face ao modelos lineares de descrever melhor determinadas características da série em estudo (tal como parece indicar também este trabalho), já no que respeita à capacidade de previsão parece não haver um consenso claro, se efectivamente a consideração de modelos não lineares conduzem a uma melhoria evidente das previsões. Quanto a este último aspecto, é importante

lembrar que efectivamente a avaliação da qualidade de previsão neste trabalho incidiu sobre os últimos três anos de observações e, portanto, apenas sobre o período durante o qual se observou o regime 2. Assim, a avaliação da qualidade de previsão que foi efectuada na prática traduz comparação do modelo linear na média condicional com o modelo não linear no regime 2. Com efeito, pode-se apenas dizer que o modelo não linear na média condicional parece revelar-se mais adequado que o linear para efeitos de previsão no regime 2, em particular, se o objectivo for o de fazer previsão trimestral<sup>51</sup>.

---

<sup>51</sup> Já quanto ao regime 1, não é possível tirar qualquer conclusão em termos de previsão *out of sample*, na medida em que como esse regime se observa apenas durante o período inicial da amostra, tal obrigaria a realizar a estimação do modelo linear e não linear apenas com o conjunto inicial de observações. O problema é que como nesse período inicial não há observações pertencentes ao segundo regime, tal impossibilita a estimação do modelo não linear.

## 4. Conclusão

Este trabalho tinha como objectivo estudar a série da taxa de juro nominal dos títulos de dívida pública *benchmark* alemães (*Bund*) com maturidade de 10 anos, procurando analisar a capacidade descritiva e de previsão de diferentes formulações lineares e não lineares na média condicional.

Em termos ajustamento aos dados *in sample* de entre as várias especificações analisadas (modelos da classe **1**, **2** e **3**) e ponderando vários factores<sup>52</sup>, os modelos que se revelaram mais satisfatórios foram os pertencentes à classe **2** (incorporação de efeitos GARCH na modelação da volatilidade). De notar que a modelação adequada da variância condicional, afim de passar os testes de diagnóstico, exigiu a inclusão de efeitos GARCH, a par da inclusão de dois regimes, o que parece confirmar a característica, frequentemente apontada às taxas de juro de longo prazo, de elevada persistência. Por sua vez, dentro dos modelos da classe **2**, quando se compara o modelo linear e não linear na média condicional, existem indícios que favorecem este último modelo face ao primeiro em termos de ajustamento aos dados históricos. Além do melhor ajustamento *in sample*, o modelo com dois regimes e efeitos GARCH na volatilidade, **2.b**)\*, permite captar determinadas características da taxa de juro de longo prazo, como o facto de os regimes parecem distinguir-se entre si não só em termos de valor médio de taxa de juro (6.7% no regime 1 e 4.4% no regime 2), como também em termos de dinâmica de volatilidade. A este respeito, no regime 1, os choques individuais parecem ter um impacto pequeno mas relativamente prolongado, enquanto no regime 2 têm um impacto imediato mais significativo mas esse efeito desaparece mais rapidamente no tempo. Adicionalmente, o regime 1 aparentemente caracteriza-se por ser um processo IGARCH, enquanto o regime 2 se define por um processo estacionário de 2º ordem.

Já em termos de previsão *out of sample* a um, dois e três meses de distância, e usando a metodologia de “janela rolante”, o modelo não linear na média condicional parece produzir resultados mais satisfatórios comparativamente ao modelo linear, acentuando a sua *performance* à medida que se alarga o horizonte de previsão. Ressalva-se, contudo, que esta avaliação foi efectuada considerando os últimos três anos

---

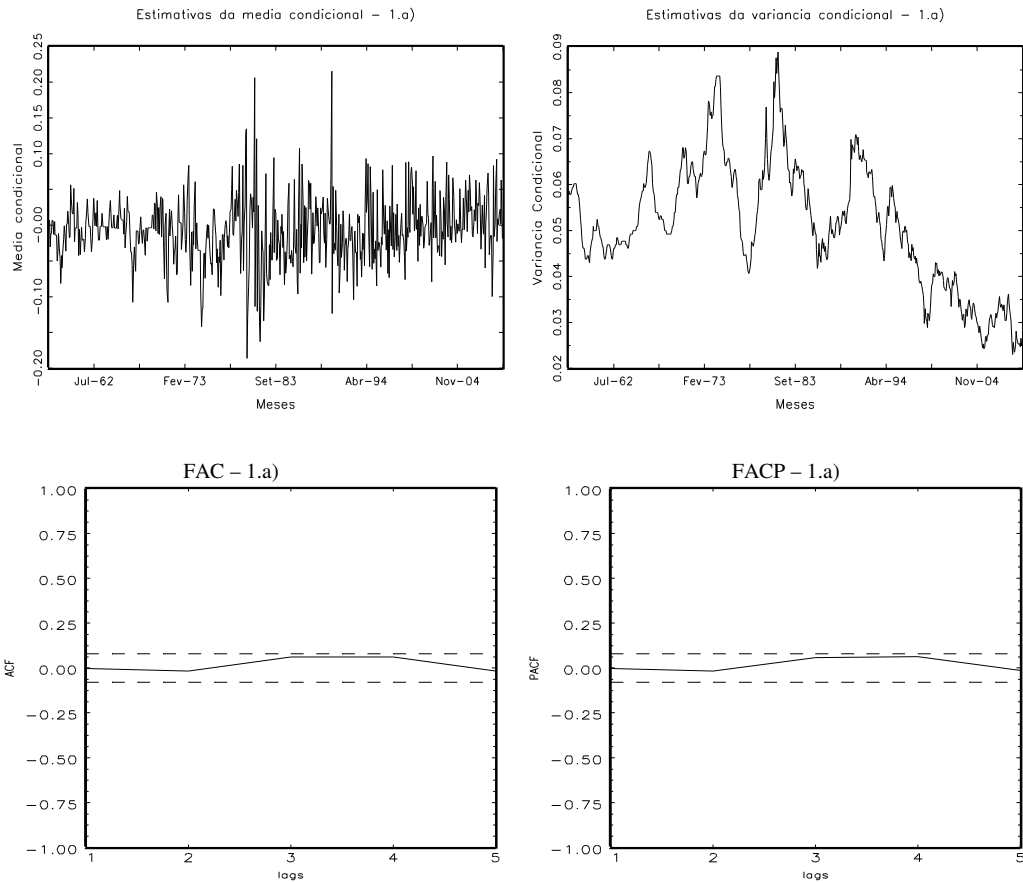
<sup>52</sup> Como sejam os critérios de informação AIC/BIC, os resultados do teste do RV e do teste *Jarque-Bera*, bem como aspectos relacionados com a robustez da estimação, entre outros.

de observações, período em que a série se encontrava com uma elevada probabilidade no regime 2. Como não foi possível efectuar uma análise análoga num período que fosse marcado por uma transição entre regimes ou com uma elevada probabilidade pelo regime 1, apenas se pode concluir que aparentemente o modelo não linear é melhor que o linear em termos de previsão quando a série se encontra com uma elevada probabilidade no segundo regime.

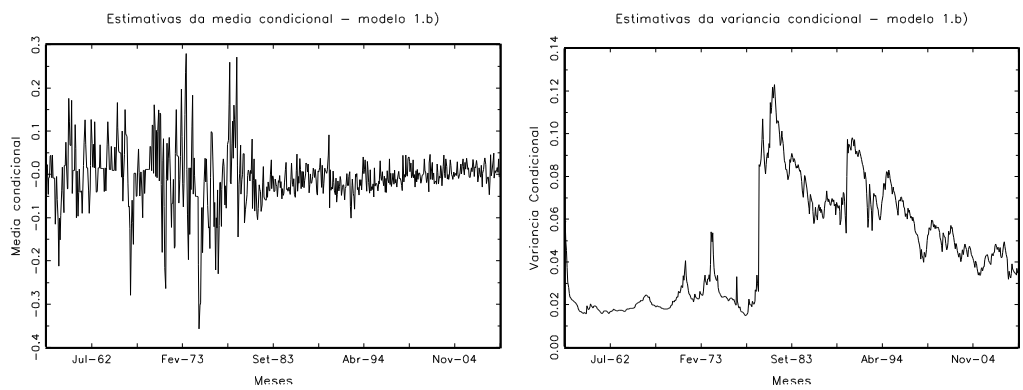
Neste sentido, quer a presença de efeitos GARCH, quer a possibilidade de alteração de regime parecem ser dois factores importantes a ter em conta na modelação da taxa de juro de longo prazo, para efeitos descritivos da série e de previsão *out of sample*.

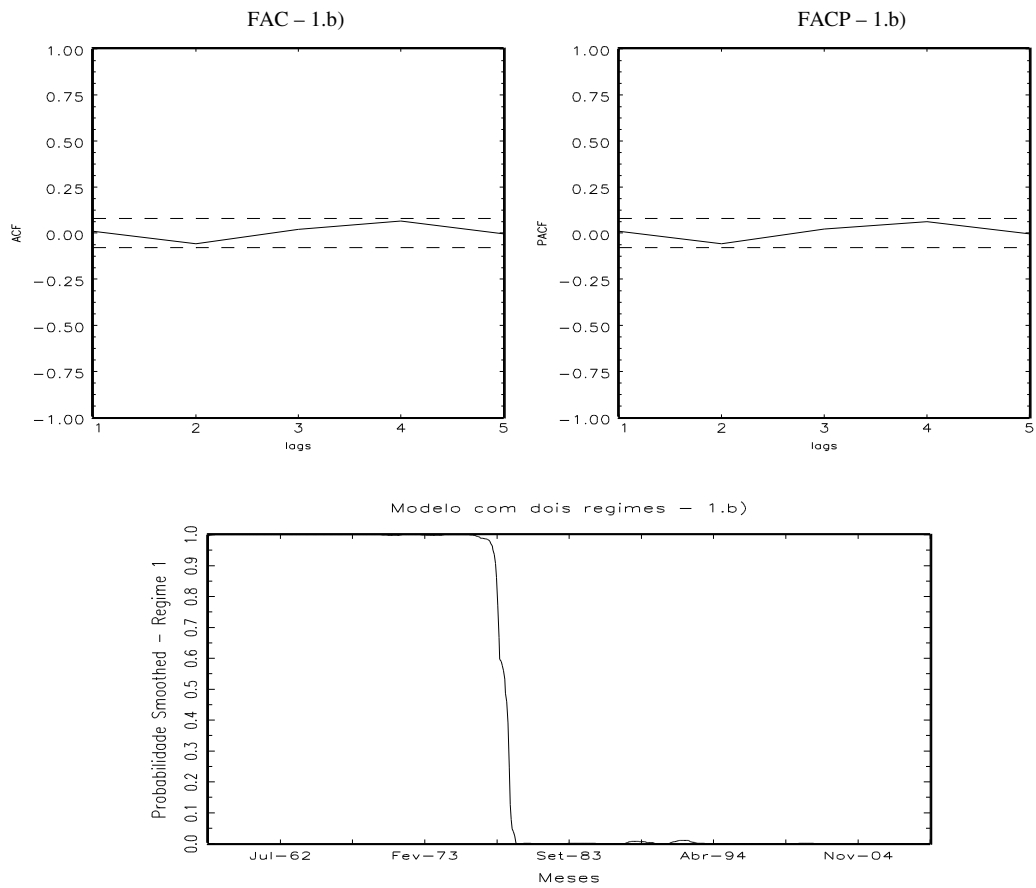
## 5. Anexo

### Modelos da classe 1: Modelos 1.a)

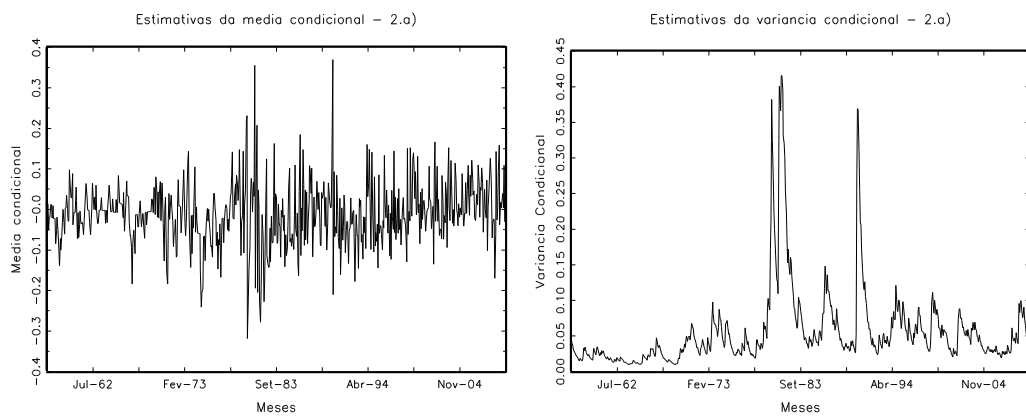


### Modelos 1.b)

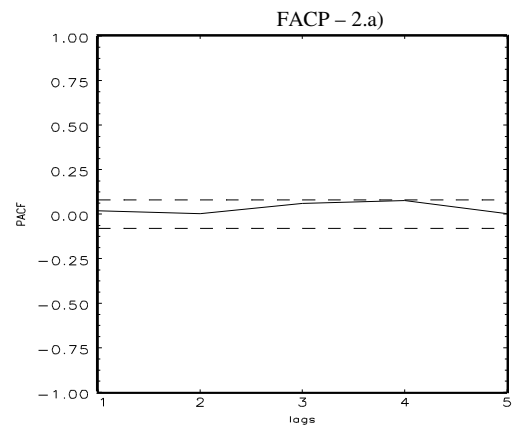
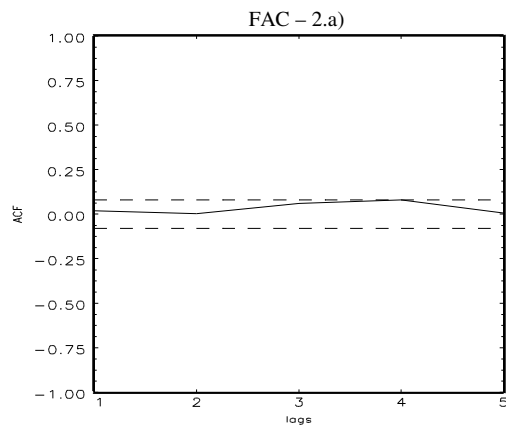




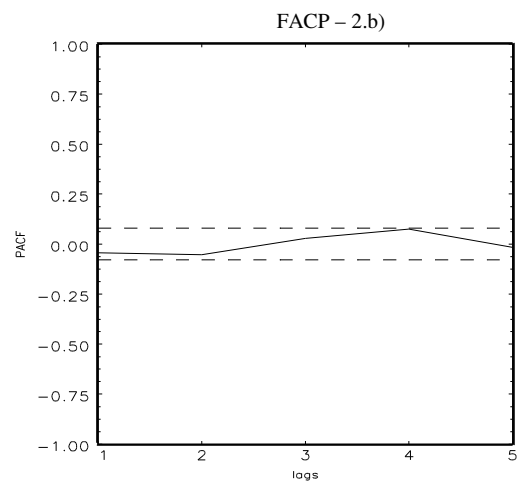
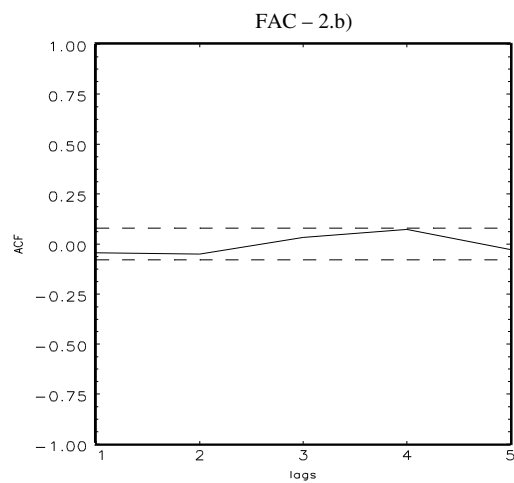
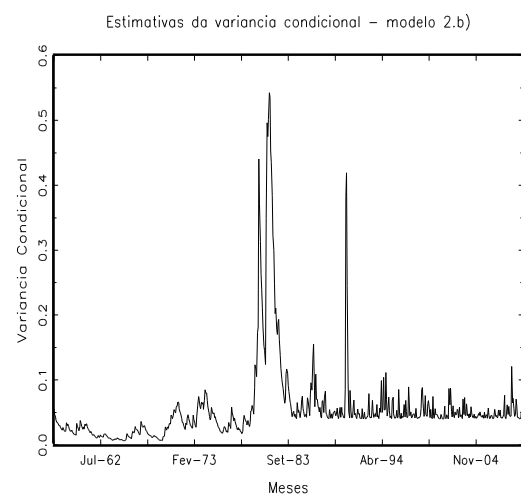
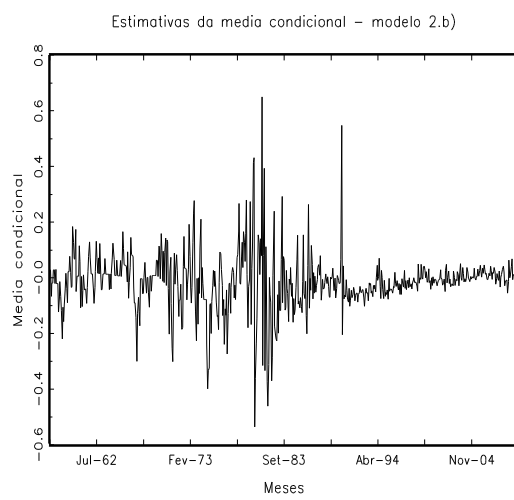
## Modelos da classe 2: Modelos 2.a)

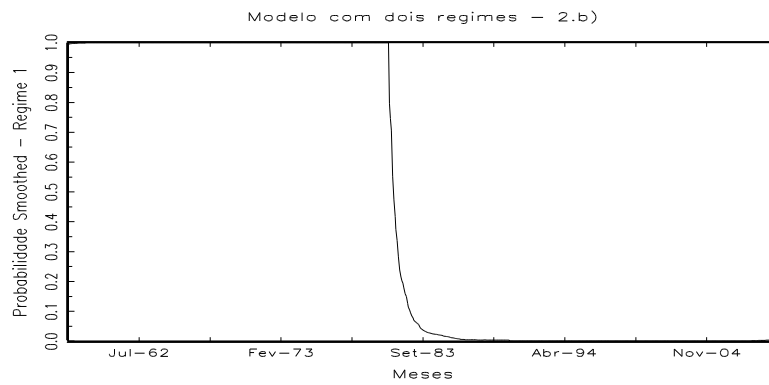






## Modelos 2.b)

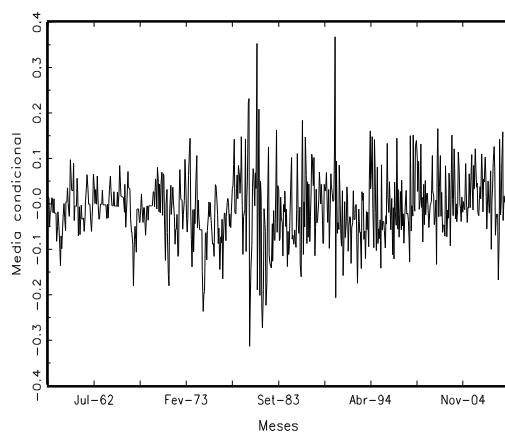




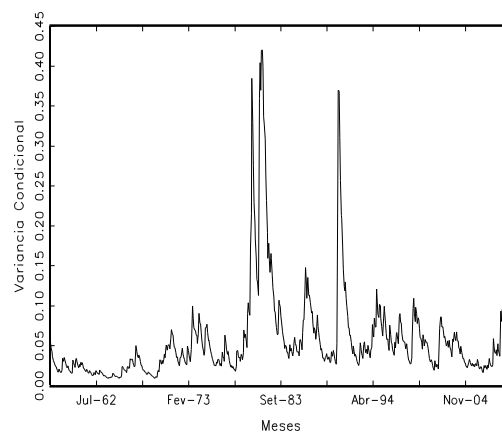
### Modelos da classe 3:

#### Modelos 3.a)

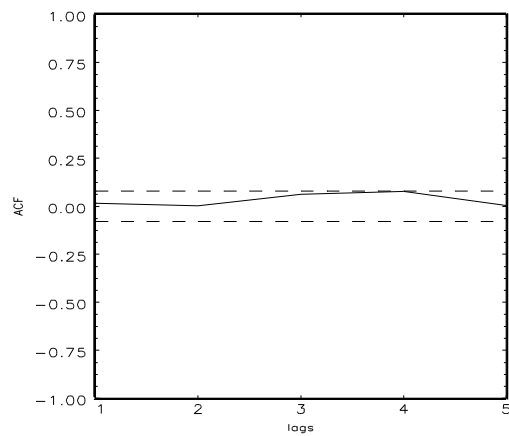
Estimativas da media condicional - 3.a)



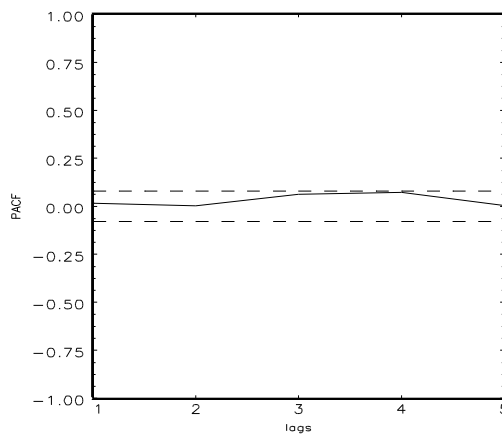
Estimativas da variancia condicional - 3.a)



FAC - 3.a)

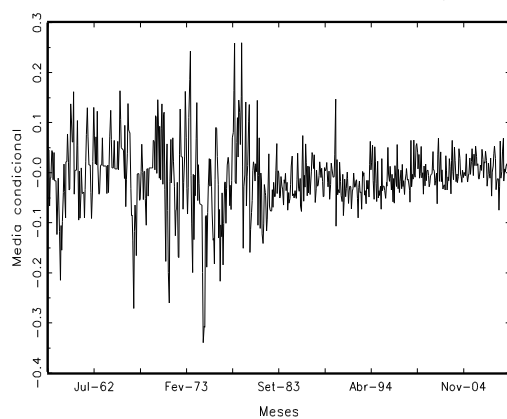


FACP - 3.a)

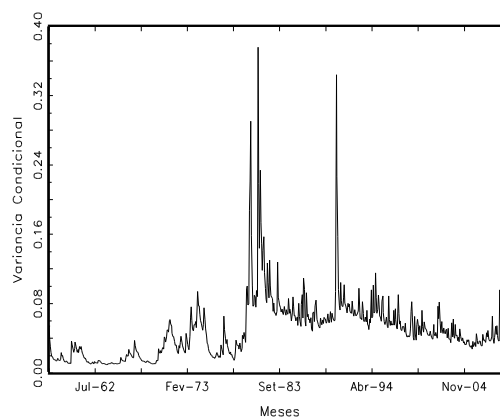


### Modelos 3.b):

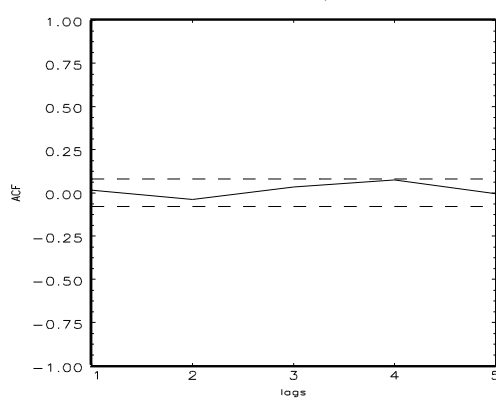
Estimativas da media condicional – modelo 3.b)



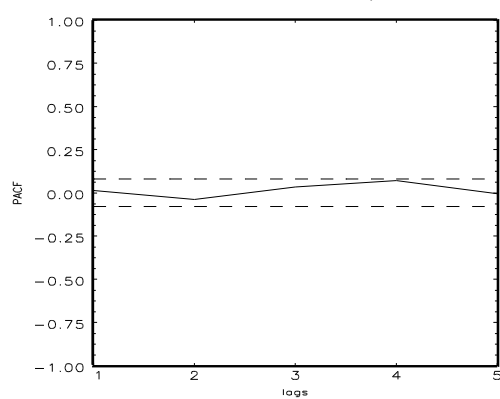
Estimativas da variancia condicional – modelo 3.b)



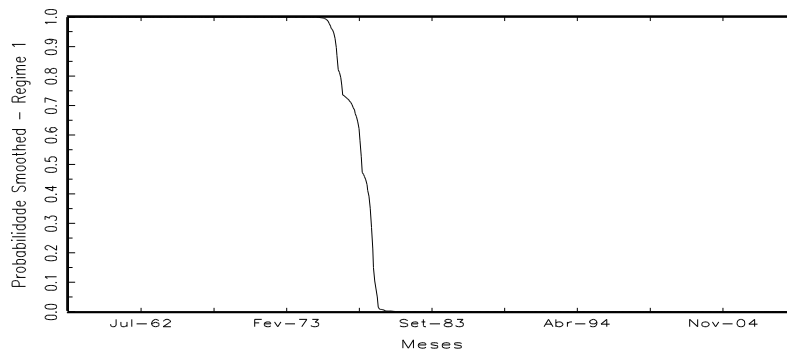
FAC – 3.b)



FACP – 3.b)



Modelo com dois regimes – 3.b)



## Bibliografia

- Ahrend R., Catta P.A. and Price R., 2006, *Factors Behind Low Long-Term Interest Rates*, OECD Working Paper No. 2006/18.
- Ang A. and Bekaert G., 2002, *Regime Switches in Interest Rate*, Journal of Business & Economic Statistics, 20, 163-182.
- Bauwens L., Preminger A. and Rombouts J., 2006, *Regime Switching GARCH Models*, Working Paper 2006006, Université catholique de Louvain, Département des Sciences Economiques.
- Bekaert G., Hodrick R. J. and Marshall D. A., 2001, *Peso Problem' Explanations for Term Structure Anomalies*, Journal of Monetary Economics, 48 (2), 241-270.
- Bollerslev Y., 1986, *Generalized autoregressive conditional heteroskedasticity*, Journal of Econometrics, 31, 307-327.
- 1988, *On the correlation structure for the generalized autoregressive conditional heteroskedastic process*, Journal of Time Series Analysis, 9, 121-131.
- Bollerslev T., Chou R.Y. and Kroner K.F., 1992, *ARCH modeling in finance*. Journal of Econometrics, 52, 5-59.
- Box G.E.P. and Jenkins G.M., 1970, *Time Series Analysis*, Forecasting and Control, San Francisco: Holden-Day.
- Cai J., 1994, *A Markov model of unconditional variance in ARCH*, Journal of Business and Economic Statistics, 12, 309-316.
- Cecchetti S. J., Lam P. and Mark N. C., 1993, *The Equity Premium and the Risk-Free Rate: Matching the Moments*, Journal of Monetary Economics, 31, 21-46.
- Chan K.C., Karolyi G.A., Longstaff F.A. and Sanders A.B., 1992, *An empirical comparison of alternative models of the short-term interest rate*, Journal of Finance, XLVII, 1209-1227.
- Chan K.C. and Tong H., 1986, *On estimating thresholds in autoregressive models*, Journal of Time Series Analysis, 7, 179-190.
- Cox J., Ingersoll J. and Ross S., 1985, *A Theory of the Term Structure of Interest Rates*, Econometrica, 53, 385-408.
- Davies, 1987, *Hypothesis testing when a nuisance parameter is present only under the alternative*, Biometrika, 74, 33-43.

- Diebold F.X., Lee J.-H., Weinbach G.C., 1994, *Regime switching with time-varying transition probabilities*, In: Hargreaves, C.P. (Ed.), *Nonstationary Time Series Analysis and Cointegration*, Oxford University Press, Oxford, pp. 283–302.
- Ding Z. and Granger C. W. J. ,1996, *Modeling volatility persistence of speculative returns: a new approach*, *Journal of Econometrics*, 73, 185-215.
- Engle R.F., 1982, *Autoregressive conditional heteroscedasticity with estimates of United Kingdom. inflation*, *Econometrica*, 50, 987-1008.
- 1983, *Estimates of the variance of US inflation based upon the ARCH model*, *Journal of Money, Credit, and Banking*, 15, 286-301.
- Engle R.F. and Bollerslev T., 1986, *Modelling the persistence of conditional variances*, *Econometric Reviews*, 5, 1-50 (with discussion).
- Filardo A.J., 1994, *Business cycle phases and their transitional dynamics*, *Journal of Business and Economic Statistics*, 12, 299–308.
- Franses P. H., Dijk D. V., 2000, *Nonlinear Time Series Models in Empirical Finance*, Cambridge University Press, Cambridge (UK).
- Garcia R. , 1998, *Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models*, *International Economic Review*, 39 (3), 763-788.
- Garcia R. and Perron P., 1996, *An Analysis of the Real Interest Rate Under Regime Shifts*, *Review of Economics and Statistics*, 78 (1), 111-125.
- Granger C.W.J., 1993, *Strategies for modelling nonlinear time-series relationships*, *The Economic Record*, 69, 233-238.
- Granger C.W.J. and Teräsvirta T., 1993, *Modelling Nonlinear Economic Relationships*, Oxford: Oxford University Press.
- Gray S. F., 1995, *An analysis of conditional regime-switching models*, Working paper (Fuqua School of Business, Duke University, Durham, NC).
- 1996, *Modeling the conditional distribution of interest rates as a regime-switching process*, *Journal of Financial Economics*, 42, 27-62.
- Hamilton J. D., 1988, *Rational expectations econometric analysis of changes in regime: An investigation of the term structure of interest rates*, *Journal of Economic Dynamics and Control*, 12, 385-423.
- 1989, *A new approach to the economic analysis of nonstationary time series and the business cycle*, *Econometrica*, 57, 357-384.
- 1990, *Analysis of time series subject to changes in regime*, *Journal of Econometrics*, 45, 39-70.
- 1994, *Time Series Models*, Princeton University Press, New Jersey.
- Hamilton J.D. and Susmel R., 1994, *Autoregressive conditional heteroscedasticity and changes in regime*, *Journal of Econometrics*, 64, 307-333.

- Hentschel L.F., 1995, *All in the family: nesting linear and nonlinear GARCH models*, Journal of Financial Economics, 39, 139-164.
- Holst U., Lindgren G., Holst J. and Thuvesholmen M., 1994, *Recursive estimation in switching autoregressions with a Markov regime*, Journal of Time Series Analysis, 15, 489-506.
- Karlin S. and Taylor H. M., 1998, *An Introduction to Stochastic Modeling*, Third Edition, Academic Press.
- Kim C. J., 1994, *Dynamic linear models with Markov-switching*, Journal of Econometrics, 60, 1-22.
- Lumsdaine R. L., 1996, *Consistency and Asymptotic Normality of the Quasi-maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models*, Econometrica, 64(3), 575-96.
- Martellini L., Priaulet P., Priaulet S., 2003, *Fixed-Income Securities Valuation, Risk Management and Portfolio Strategies*, Wiley, England.
- Nelson D. B., 1990, *Stationarity and persistence in the GARCH(1,1) model*, Econometric Theory, 6, 318-334.
- Nicolau J., 2010, *Econometria Financeira*, disponível em [http://pascal.iseg.utl.pt/~nicolau/myHP/manual\\_EF.pdf](http://pascal.iseg.utl.pt/~nicolau/myHP/manual_EF.pdf)
- Stelzer R., 2009, *On Markov-Switching Arma Processes? Stationarity, Existence Of Moments, And Geometric Ergodicity*, Econometric Theory, Cambridge University Press, 25(01), 43-62.
- Taylor S. J., 1986, *Modelling Financial Time Series*, New York: John Wiley.
- Teräsvirta T., 1994, *Specification, estimation, and evaluation of smooth transition autoregressive models*, Journal of the American Statistical Association, 89, 208-218.
- 1998, *Modelling economic relationships with smooth transition regressions*, in A. Ullah and D.E.A. Giles (eds.), Handbook of Applied Economic Statistics, New York: Marcel Dekker, 507-552.
- Tong H., 1978, *On a threshold model*, in C.H. Chen (ed.), Pattern Recognition and Signal Processing, Amsterdam: Sijthoff & Noordhoff, 101-141.
- 1990, *Non-Linear Time Series: A Dynamical Systems Approach*, Oxford: Oxford University Press.
- Tong H. and Lim K.S., 1980, *Threshold autoregressions, limit cycles, and data*, Journal of the Royal Statistical Society B 42, 245-292 (with discussion).
- White H., 1980, *A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity*, Econometrica, 48, 817-838.